



UNIVERSIDAD DE LA RIOJA

TESIS DOCTORAL

Título
Lexical availability and motivation in CLIL: the effects of language of instruction and gender
Autor/es
Leah Geoghegan
Director/es
María del Pilar Agustín Llach
Facultad
Facultad de Letras y de la Educación
Titulación
Departamento
Filologías Modernas
Curso Académico



Lexical availability and motivation in CLIL: the effects of language of instruction and gender, tesis doctoral de Leah Geoghegan, dirigida por María del Pilar Agustín Llach (publicada por la Universidad de La Rioja), se difunde bajo una Licencia Creative Commons Reconocimiento-NoComercial-SinObraDerivada 3.0 Unported. Permisos que vayan más allá de lo cubierto por esta licencia pueden solicitarse a los titulares del copyright.



**UNIVERSIDAD
DE LA RIOJA**

TESIS DOCTORAL

2022

**LEXICAL AVAILABILITY AND MOTIVATION IN
CLIL: THE EFFECTS OF LANGUAGE OF
INSTRUCTION AND GENDER**

**DISPONIBILIDAD LÉXICA Y MOTIVACIÓN EN
AICLE: LOS EFECTOS DEL IDIOMA DE
INSTRUCCIÓN Y GÉNERO**

Programa de Doctorado en Filología Inglesa

Leah Geoghegan

Directora: M^a Pilar Agustín Llach

Acknowledgements

I would like to acknowledge the help and support of the following people, without whom this doctoral thesis would not have seen the light of day and to whom I am eternally grateful.

My most heartfelt gratitude goes to my doctoral supervisor, M^a Pilar Agustín Llach, whose constant support and dedication have made this work possible. With her invaluable guidance, moral support, and kind encouragement, she has gone above and beyond the call of duty and has truly made this experience not only one of the most rewarding, but also one of the most enjoyable ones of my life.

I would like to thank the members of the GLAUR research group, from whom I have learnt a great deal and whose work has inspired and continues to inspire me along this journey.

My appreciation also goes to Julio Rubio García for his assistance in preparing the raw data. His expertise saved countless hours of work in this challenging part of the process.

I would like to acknowledge the kind involvement of the participating schools, without whom this project could not have gone ahead. I would especially like to convey my gratitude to the programme coordinators and teachers who, despite their busy schedules, were incredibly generous and welcoming with their time and always willing to provide assistance whenever necessary.

On a personal note, I would like to thank my family, and in particular my husband. His unconditional support and words of encouragement have always helped me to persevere throughout my early research career and the entire process of completing this PhD. I cannot thank him enough for his guidance, love and support and can truly say that, without him, this doctoral thesis would not have been possible.

Abstract

Over the last two decades, Content and Language Integrated Learning (CLIL) has become widely implemented throughout Spain, with the hope of enhancing foreign-language abilities and encouraging multilingualism (Ruiz de Zarobe & Lasagabaster, 2010). However, there has been such enthusiasm for the approach that its implementation has largely outpaced research into its effectiveness (Pérez-Cañado, 2012). In particular, there are four key issues which need to be addressed: the language of instruction, the acquisition of vocabulary, the learners' motivation, and the influence of gender.

Firstly, the vast majority of CLIL programmes throughout Spain are taught through the medium of English, prompting the use of the term CEIL (content-and-English integrated learning) (Dalton-Puffer, 2011). However, several schools also implement programmes through other important target languages (TLs) such as French (Ruiz de Zarobe & Lasagabaster, 2010). These languages other than English, however, have largely been neglected, both in research and practice. Within CLIL research Dalton-Puffer, Nikula and Smit (2010) have consequently called for comparative research across different TLs, so as to provide insight into the strengths and weaknesses of CLIL language-independently. Secondly, CLIL modules have been shown to produce positive effects on students' learning of content-related vocabulary (Heras & Lasagabaster, 2015). Within the field of lexical availability (LA), there have been recent calls to better understand this effect, by focusing on LA prompts which may be relevant to the CLIL subject and by including some measure of proficiency in order to determine its influence (Canga Alonso, 2017). Thirdly, motivation is well understood as an undeniably important factor in learning a foreign language (FL) and has been found to play a more significant role in CLIL than in non-CLIL settings (Navarro & García Jiménez, 2018). However, there may be a fundamental difference between the motivation for learning English compared with other languages, given that English is increasingly regarded as a basic educational skill, crucial to professional development (Dörnyei & Ushioda, 2013). There is thus a clear need to investigate CLIL language learning motivation in English as compared to other TLs, so as to determine whether the benefits hold true for languages other than English. Finally, research into gender and LA in Spain has largely shown a female advantage in younger learners and in early secondary school students. However, more research is needed to understand whether this advantage remains in later school

years, and whether CLIL instruction plays a role. Female students have also generally been found to exhibit higher language learning motivation, however, there are suggestions that a CLIL context may provide a blurring effect of gender differences (Lasagabaster, 2008), given the assumption that male students may compensate for lower FL learning motivation with higher motivation towards the CLIL subject (Heras & Lasagabaster, 2015). However, research to date has produced mixed findings on whether this is in fact the case.

To address these gaps, this thesis aims to analyse the LA and language learning motivation of secondary school students enrolled simultaneously in English and French CLIL, exploring the influence of the language of instruction, the role of gender, and the effect of CLIL on these different factors. The participants are native Spanish speakers from 9th, 10th and 11th grade who in addition to studying English as a Foreign Language (EFL) and French as a Foreign Language (FFL) also study content subjects through the medium of these languages. The study adopts both a cross-sectional (9th to 10th grade) and longitudinal (10th to 11th grade) approach, with participants completing language level C-tests, LA tasks and language learning motivation questionnaires both in English and French.

Results point to clear differences between participants' LA and language learning motivation in English and French, to the advantage of English. In terms of LA, exposure to content-related vocabulary was found to influence the results from one grade to the next, and language level was found to play a clear role. Regarding motivation, in addition to reporting greater motivation towards English, there was a much stronger relationship between LA and motivation and between language level and motivation in English than in French. There were also indications that attitudes toward CLIL classes may play a vital role. In terms of gender, female students were found to produce a higher number of words than male students, however, there are indications that male students may exhibit greater lexical sophistication in terms of LA. While male and female students were equally motivated towards learning English, clearer differences arose in French. Finally, CLIL instruction was also found to play a very important role in the acquisition of content-related vocabulary. In particular, results revealed that CLIL has the potential to help students improve their LA to the extent that they can effectively bridge the gap between them and students with a higher language level and greater LA in other domains.

Resumen

En las últimas dos décadas, el Aprendizaje Integrado de Contenidos y Lenguas Extranjeras (AICLE) ha crecido a gran escala por toda España, con el fin de mejorar los conocimientos de los idiomas extranjeros y fomentar el multilingüismo (Ruiz de Zarobe y Lasagabaster, 2010). Sin embargo, ha habido tanto entusiasmo que su implementación ha adelantado en gran medida la investigación sobre su eficacia (Pérez-Cañado, 2012). En concreto, es preciso abordar cuatro asuntos claves: el idioma de instrucción, la adquisición de vocabulario, la motivación del alumno, y la influencia del género.

En primer lugar, la gran mayoría de los programas AICLE en España se imparten en inglés, lo que ha hecho surgir el concepto de Aprendizaje Integrado de Contenidos en Inglés (CEIL; por sus siglas en inglés) (Dalton-Puffer, 2011). No obstante, también existen programas en otras lenguas europeas, como el francés (Ruiz de Zarobe y Lasagabaster, 2010). Estos idiomas distintos del inglés, sin embargo, se han desatendido en gran medida, tanto en la investigación como en la práctica. Por consiguiente, dentro de la investigación AICLE, Dalton-Puffer, Nikula y Smit (2010) han hecho un llamamiento a la investigación comparativa de diferentes lenguas meta, con el fin de comprender las fortalezas y debilidades de AICLE independientemente de la lengua del programa. En segundo lugar, se ha demostrado que la instrucción AICLE tiene efectos positivos en el aprendizaje del vocabulario relacionado con el contenido (Heras y Lasagabaster, 2015). Dentro del campo de la disponibilidad léxica (DL), ha habido llamadas para entender este efecto, y en concreto para centrarse en los centros de interés que están relacionados con la propia asignatura AICLE e incluir alguna medida de aptitud lingüística para determinar su influencia (Canga Alonso, 2017). En tercer lugar, es bien sabido que la motivación es un factor de innegable importancia en el aprendizaje de un idioma extranjero, y se ha demostrado que desempeña un papel más significativo en un contexto AICLE que no-AICLE (Navarro y García Jiménez, 2018). No obstante, puede que haya una diferencia fundamental entre la motivación para aprender el inglés frente a otras lenguas, dado que el inglés cada vez más se ve como una habilidad educativa básica, crucial para el desarrollo profesional (Dörnyei y Ushioda, 2013). Por eso, es necesario investigar la motivación en AICLE hacia el inglés en comparación a otras lenguas meta, para determinar si los beneficios siguen siendo válidos en otros idiomas. En último lugar, la investigación sobre género y DL ha comprobado que hay una ventaja

femenina en los últimos años de la escuela primaria y los primeros de la secundaria. Sin embargo, se necesita más investigación para entender si tal ventaja continua en las siguientes etapas escolares y si la instrucción AICLE contribuye a esta ventaja. También se ha comprobado que las alumnas están más motivadas hacia las lenguas extranjeras, no obstante, se ha sugerido que un contexto AICLE tiene la capacidad de desdibujar las diferencias de género (Lasagabaster, 2008) dada la suposición que los chicos pueden compensar su menor motivación hacia la lengua extranjera con una mayor motivación hacia la asignatura AICLE (Heras y Lasagabaster, 2015). No obstante, la investigación previa ha producido hallazgos contradictorios sobre si este es el caso.

Para abordar estos vacíos, esta tesis pretende analizar la DL y la motivación hacia las lenguas extranjeras de alumnos de secundaria, que cursan a la vez asignaturas de instrucción AICLE en inglés y francés, y explorar la influencia del idioma de instrucción, el papel del género, y el efecto de AICLE en estos diferentes factores. Los participantes son españoles nativos de 9º, 10º y 11º curso (3º ESO, 4º ESO y 1º Bachillerato), que además de estudiar inglés y francés como lenguas extranjeras, también cursan asignaturas AICLE en inglés y francés. El estudio adopta un enfoque transversal (9º a 10 grado) y longitudinal (10º a 11º), y los participantes realizan pruebas de DL, cuestionarios sobre la motivación y pruebas C-test, tanto en inglés como en francés.

Los resultados indican diferencias claras entre la DL y motivación de los participantes en inglés y francés, con ventaja del inglés. Respecto a la DL, se ha demostrado que la exposición a vocabulario relacionado con el contenido de la clase influyó en los resultados de un grado al siguiente, y que la competencia lingüística tuvo un papel importante. Respecto a la motivación, además de encontrar un nivel más alto de motivación hacia el inglés, hubo una relación mucho más fuerte entre la DL y la motivación y entre el nivel de idioma y la motivación en inglés que en francés. También se indicó que las actitudes hacia las clases AICLE pueden ser de suma importancia. Respecto al género, se constató que las chicas produjeron más palabras que los chicos, sin embargo, hay evidencia que los chicos demuestran una mayor sofisticación léxica en su DL. Mientras los chicos y las chicas indicaron un nivel de motivación parecido hacia el inglés, había más diferencias en su motivación hacia el francés. Por último, se comprobó que la instrucción AICLE juega un importante papel en la adquisición de vocabulario relacionado con el contenido de la clase. En concreto, los resultados

revelaron que la instrucción AICLE puede ayudar a que los alumnos mejoren su disponibilidad hasta efectivamente romper la distancia entre ellos y otros alumnos con un nivel de idioma más alto y que producen más vocabulario en otros ámbitos lingüísticos.

Table of Contents

List of Tables.....	xiii
List of Figures.....	xix
List of Abbreviations.....	xxi
Chapter 1: Introduction.....	23
Chapter 2: Vocabulary and Second Language Acquisition.....	29
2.1. The Importance of Vocabulary in Second Language Acquisition.....	29
2.2. Challenges in Learning Vocabulary	30
2.3. Vocabulary Research: Dimensions of Word Knowledge	33
2.3.1. Lexical Competence.....	33
2.3.2. Dimensions in lexical knowledge	35
2.4. Vocabulary Research: Measurement	41
2.4.1. Measuring Receptive Knowledge	45
2.4.1.1. The VLT/VST Approach.....	45
2.4.1.2. The Yes/No Test Approach.....	47
2.4.1.3. The Translation Approach.....	48
2.4.2. Measuring Productive Knowledge.....	50
2.4.2.1. Translation and Elicitation	51
2.4.2.2. Statistical Analysis of Free Production.....	54
2.4.2.2.1. Lexical Diversity.....	54
2.4.2.2.2. Lexical Sophistication.....	55
2.4.2.2.3. Lexical Density	58
2.4.2.2.4. Proportion of Errors	59
2.4.2.3. Association Tests.....	60
2.4.2.4. Measures of Automaticity	63
2.5. Lexical Availability	65
2.5.1. An Overview of Lexical Availability.....	65

2.5.2. The Lexical Availability Task.....	67
2.5.3. Previous Research on Lexical Availability	68
Chapter 3: Motivation in Second Language Learning	77
3.1. Motivation in Second Language Acquisition	77
3.1.1. What is L2 Motivation and Why is it Important?	77
3.1.2. Key Terms in L2 Motivation.....	78
3.2. An Overview of L2 Learning Motivation: Key Theories and Approaches	82
3.2.1. The Social-Psychological Period	83
3.2.2. The Cognitive-Situated Period	89
3.2.3. The Process-Oriented Period.....	94
3.2.4. The Socio-Dynamic Period	97
3.3. L2 Motivation Research.....	103
3.4. Research on Motivation in English and LOTEs	105
3.4.1. Language Motivation in English and LOTEs	105
3.4.2. Language Motivation Research on English and LOTEs.....	107
3.4.2.1. Research on Motivation towards LOTEs.	108
3.4.2.2. Research Comparing Motivation towards English and LOTEs.	110
3.5. Research on Motivation and the Learning Context	119
3.5.1. L2 Motivation in Formal Instruction.....	119
3.5.2. L2 Motivation in Study Abroad	127
3.5.3. L2 Motivation in CLIL.....	133
3.5.3.1. Measurement of Motivation in CLIL.	134
3.5.3.1.1. Measurement of Motivation in CLIL in Spain.	134
3.5.3.1.2. Measurement of Motivation in CLIL in Europe.....	140
3.5.3.2. Stakeholders' beliefs in CLIL.....	146
3.6. Research on Motivation and Gender	150

Chapter 4: Gender.....	157
4.1. Gender and Language Acquisition	157
4.2. Gender in Second Language Acquisition	158
4.3. Research on Gender and Vocabulary, Motivation and CLIL	161
4.3.1. Gender and Vocabulary.....	161
4.3.2. Gender and Motivation.....	170
4.3.3. Gender and CLIL	171
Chapter 5: Content and Language Integrated Learning.....	177
5.1. An Overview of Content and Language Integrated Learning.....	177
5.1.1. What is Content and Language Integrated Learning?	177
5.1.1.1. CLIL, CBI or Immersion	178
5.1.1.2. CLIL versus EMI.....	179
5.1.2. Advantages and Disadvantages of CLIL.....	180
5.1.2.1. Reasons for adopting a CLIL approach.	180
5.1.2.2. Potential shortcomings of a CLIL approach.....	184
5.1.2.2.1. Shortcomings in CLIL research	184
5.1.2.2.2. Issues in the content and language balance	186
5.2. The Context.....	192
5.2.1. CLIL in Spain.....	192
5.2.2. CLIL in La Rioja	197
5.2.2.1. School Language Innovation Projects	197
5.2.2.2. School Integrated Language Competence Programme.	199
5.2.2.3. Bilingual Sections.....	199
5.3. Research on Vocabulary and CLIL	200
5.3.1. Receptive Vocabulary in CLIL	202
5.3.2. Productive Vocabulary in CLIL	209

5.4. Research on Motivation and CLIL	219
5.5. Research on Gender and CLIL	220
Chapter 6: Research Questions and Hypotheses	223
6.1. Research Rationale and Objectives.....	223
6.2. Research Questions and Hypotheses	227
6.2.1. Target Language Influence on Lexical Availability	227
6.2.2. Target Language Influence on Language Learning Motivation.....	228
6.2.3. Gender	229
6.2.4. Content and Language Integrated Learning	231
Chapter 7: Methodology.....	233
7.1. Research Approach and Design	233
7.2. Participants.....	235
7.3. Data collection	238
7.3.1. Instruments	238
7.3.1.1. C-tests	238
7.3.1.2. Lexical Availability Tasks.....	238
7.3.1.3. Questionnaires	239
7.3.2. Procedure.....	243
7.4. Data Analysis	244
7.4.1. C-tests.....	244
7.4.2. Lexical Availability Tasks	244
7.4.3. Questionnaires	248
Chapter 8: Results.....	251
8.1. Target Language Influence on Lexical Availability	251
8.1.1. Quantitative Differences in English and French LA	251
8.1.2. Qualitative Differences in English and French LA	261
8.2. Target Language Influence on Language Learning Motivation	271

8.2.1. Quantitative Differences in Motivation towards English and French.....	271
8.2.2. Lexical Availability, Language Level and Language Learning Motivation	280
8.3. Gender.....	284
8.3.1. Quantitative Differences in Gender and LA	284
8.3.1.1. Quantitative Differences in Gender and LA across Languages	284
8.3.1.2. Quantitative Differences in Gender and LA within Languages.	305
8.3.2. Qualitative Differences in Gender and LA	326
8.3.3. Quantitative Differences in Gender and Motivation	341
8.3.3.1. Quantitative Differences in Gender and Motivation across Languages.	342
8.3.3.2. Quantitative Differences in Gender and Motivation within Languages.	354
8.4. Content and Language Integrated Learning.....	372
8.4.1. Quantitative Differences in CLIL: Language Level, Lexical Availability and Motivation	373
8.4.2. Qualitative Differences in CLIL: Lexical Availability	384
8.4.3. Quantitative Differences in CLIL: Content-Relevant Prompts.....	389
8.4.4. Qualitative Differences in CLIL: Content-Relevant Prompts.....	392
Chapter 9: Discussion.....	395
9.1. Target Language Influence on Lexical Availability	395
9.1.1. Quantitative Differences in English and French LA.....	395
9.1.2. Qualitative Differences in English and French LA	397
9.2. Target Language Influence on Language Learning Motivation	402
9.2.1. Quantitative Differences in Motivation towards English and French.....	402
9.2.2. Lexical Availability, Language Level and Language Learning Motivation	404
9.3. Gender.....	406
9.3.1. Quantitative Differences in Gender and LA	406

9.3.1.1. Quantitative Differences in Gender and LA across Languages	406
9.3.1.2. Quantitative Differences in Gender and LA within Languages	408
9.3.2. Qualitative Differences in Gender and LA	411
9.3.3. Quantitative Differences in Gender and Motivation	413
9.3.3.1. Quantitative Differences in Gender and Motivation across Languages	414
9.3.3.2. Quantitative Differences in Gender and Motivation within Languages	417
9.4. Content and Language Integrated Learning.....	419
9.4.1. Quantitative Differences in CLIL: Language Level, Lexical Availability and Motivation	420
9.4.2. Qualitative Differences in CLIL: Lexical Availability	422
9.4.3. Quantitative Differences in CLIL: Content-Relevant Prompts.....	424
9.4.4. Qualitative Differences in CLIL: Content-Relevant Prompts.....	425
Chapter 10: Summary and Conclusions	429
10.1. Conclusions.....	429
10.2. Limitations	432
10.3. Directions for Further Research.....	435
References	437
Appendices	499
Appendix A: English C-test.....	500
Appendix B: French C-test	502
Appendix C: English Lexical Availability Task.....	504
Appendix D: French Lexical Availability Task.....	510
Appendix E: English Language Learning Motivation Questionnaire.....	516
Appendix F: French Language Learning Motivation Questionnaire.....	519
Appendix G: School Consent Form.....	522
Appendix H: Student Consent Form.....	522
Appendix I: Conclusiones.....	523

List of Tables

Table 2.1 <i>Distinctions in Lexical Knowledge</i>	35
Table 2.2 <i>Differences between Receptive and Productive Knowledge</i>	36
Table 2.3 <i>What is Involved in Knowing a Word</i>	37
Table 2.4 <i>Dimensions of Vocabulary Assessment</i>	40
Table 2.5 <i>Summary of Bauer and Nation’s (1993) Word Families Scheme</i>	43
Table 2.6 <i>Key Foreign Language LA Studies in Spain</i>	69
Table 3.1 <i>Motivational Dichotomies</i>	81
Table 3.2 <i>Scales Contributing to the Affective Constructs in the Socio-Educational Model</i>	86
Table 3.3 <i>Research on Motivation towards LOTEs</i>	108
Table 3.4 <i>Research Comparing Motivation towards English and LOTEs</i>	110
Table 3.5 <i>Research on L2 Motivation in FI</i>	120
Table 3.6 <i>Research on L2 Motivation in Study Abroad</i>	128
Table 3.7 <i>Quantitative and Qualitative Measurement of Motivation in CLIL in Spain</i>	135
Table 3.8 <i>Quantitative and Qualitative Measurement of Motivation in CLIL in Europe</i>	141
Table 3.9 <i>Motivation Research on Stakeholders’ Beliefs in CLIL</i>	147
Table 3.10 <i>Research on Gender and Motivation</i>	151
Table 4.1 <i>Research on Gender and Vocabulary in a Spanish Context</i>	163
Table 4.2 <i>Research on Gender in CLIL</i>	172
Table 5.1 <i>Language Competencies Favourably Affected or Unaffected by CLIL</i>	182
Table 5.2 <i>Types of CLIL Provision in Europe</i>	191
Table 5.3 <i>CLIL Instruction through Different Languages in Spain</i>	193
Table 5.4 <i>CLIL Programmes in Spain</i>	195
Table 5.5 <i>School Language Innovation Projects Modalities</i>	198
Table 5.6 <i>Summary of the School Language Innovation Project in the 2019/2020 and 2020/2021 Academic Years</i>	198
Table 5.7 <i>Summary of Research Investigating Receptive Vocabulary in Spanish CLIL</i>	203
Table 5.8 <i>Research Investigating Productive Vocabulary in Spanish CLIL: Vocabulary Tests</i>	210

Table 5.9 <i>Research Investigating Productive Vocabulary in Spanish CLIL: Written Compositions</i>	213
Table 7.1 <i>Longitudinal and Cross-Sectional Approach of the Study</i>	233
Table 7.2 <i>Design of the Study</i>	234
Table 7.3 <i>Participants in the Study</i>	235
Table 7.4 <i>Participants' CLIL Subjects</i>	236
Table 7.5 <i>Subgroups of Students Taking Different CLIL subjects in 10th and 11th Grade</i>	237
Table 7.6 <i>Lexical Availability Task Prompts</i>	239
Table 7.7 <i>Sample MFQ Question with Five-Level Likert Scale Format</i>	240
Table 7.8 <i>Changes to the Category "Attitude to Learning: Past and Present"</i>	241
Table 7.9 <i>Additional Questions in the Category "Attitude to Learning: Past and Present"</i>	242
Table 7.10 <i>Categorisation of Excluded Words in the LAT</i>	245
Table 8.1 <i>Descriptive Statistics for Lexical Availability in English and French</i>	252
Table 8.2 <i>Normality Tests for Lexical Availability in English and French</i>	253
Table 8.3 <i>Differences in Lexical Availability in English and French</i>	254
Table 8.4 <i>Cross-Sectional Differences in the English and French LATs</i>	256
Table 8.5 <i>Longitudinal Differences in the English and French LATs</i>	258
Table 8.6 <i>Normality Tests for C-tests</i>	259
Table 8.7 <i>Most Frequent First Words for Each Prompt</i>	262
Table 8.8 <i>Ranking of Most and Least Productive Prompts</i>	263
Table 8.9 <i>Cognates in the LAT</i>	264
Table 8.10 <i>Non-shared Words</i>	265
Table 8.11 <i>Frequency Distributions English</i>	267
Table 8.12 <i>Frequency Distributions French</i>	268
Table 8.13 <i>Descriptive Statistics for Language Learning Motivation in English and French</i>	272
Table 8.14 <i>Results of Wilcoxon Signed-rank Tests Comparing MFQ in English and French</i>	273
Table 8.15 <i>Cross-Sectional Differences in English Language Learning Motivation</i> ..	276
Table 8.16 <i>Cross-Sectional Differences in French Language Learning Motivation</i> ...	276

Table 8.17 <i>Longitudinal Differences in Longitudinal English Language Learning Motivation</i>	278
Table 8.18 <i>Longitudinal Differences in French Language Learning Motivation</i>	278
Table 8.19 <i>Spearman Rank Correlations between Mean LAT and MFQ Categories in English</i>	281
Table 8.20 <i>Spearman Rank Correlations between LAT and MFQ Categories in French</i>	282
Table 8.21 <i>Summary of Non-significant Correlations in English and French</i>	284
Table 8.22 <i>Descriptive Statistics for Lexical Availability within Gender across Languages</i>	285
Table 8.23 <i>Normality Tests in LAT for Gender</i>	287
Table 8.24 <i>Differences in Lexical Availability in English and French by Gender</i>	290
Table 8.25 <i>Cross-Sectional Differences for Male Participants in the English and French LATs</i>	295
Table 8.26 <i>Cross-Sectional Differences for Female Participants in the English and French LATs</i>	296
Table 8.27 <i>Longitudinal Differences for Male Participants in the English and French LATs</i>	299
Table 8.28 <i>Longitudinal Differences for Female Participants in the English and French LATs</i>	300
Table 8.29 <i>Summary of Cross-Sectional and Longitudinal Differences in English and French C-tests and LATs by Gender</i>	302
Table 8.30 <i>Descriptive Statistics for Lexical Availability by Gender in English</i>	306
Table 8.31 <i>Differences in Lexical Availability in English by Gender</i>	307
Table 8.32 <i>Normality Tests for the English C-test for Gender</i>	308
Table 8.33 <i>Descriptive Statistics for Lexical Availability by Gender in French</i>	310
Table 8.34 <i>Differences in Lexical Availability in French by Gender</i>	311
Table 8.35 <i>Normality Tests for the French C-test for Gender</i>	312
Table 8.36 <i>Summary of Statistically Significant Differences between Male and Female Participants within Languages</i>	313
Table 8.37 <i>Cross-Sectional Differences in the English LAT by Gender</i>	316
Table 8.38 <i>Cross-Sectional Differences in the French LAT by Gender</i>	317

Table 8.39 <i>Longitudinal Differences in the English LAT by Gender</i>	320
Table 8.40 <i>Longitudinal Differences in the French LAT by Gender</i>	321
Table 8.41 <i>Summary of Cross-Sectional Differences in LAT and C-test by Gender within Languages</i>	322
Table 8.42 <i>Summary of Longitudinal Differences in LAT and C-test by Gender within Languages</i>	324
Table 8.43 <i>Most Frequent First Words in English by Gender</i>	327
Table 8.44 <i>Most Frequent First Words in French by Gender</i>	328
Table 8.45 <i>Ranking of Most and Least Productive Prompts in English by Gender</i>	330
Table 8.46 <i>Ranking of Most and Least Productive Prompts in French by Gender</i>	331
Table 8.47 <i>Cognates in the English LAT by Gender</i>	332
Table 8.48 <i>Cognates in the French LAT by Gender</i>	333
Table 8.49 <i>Non-shared Words in English by Gender</i>	335
Table 8.50 <i>Non-shared Words in French by Gender</i>	336
Table 8.51 <i>Frequency Distributions English by Gender</i>	339
Table 8.52 <i>Frequency Distributions French by Gender</i>	340
Table 8.53 <i>Descriptive Statistics for Language Learning Motivation within Gender across Languages</i>	342
Table 8.54 <i>Differences in Language Learning Motivation within Gender across Languages</i>	344
Table 8.55 <i>Summary of Quantitative Differences in Gender and Motivation across Languages</i>	347
Table 8.56 <i>Cross-Sectional Differences in Language Learning Motivation by Male Participants in English and French</i>	350
Table 8.57 <i>Cross-Sectional Differences in Language Learning Motivation by Female Participants in English and French</i>	351
Table 8.58 <i>Longitudinal Differences in Language Learning Motivation by Male Participants in English and French</i>	352
Table 8.59 <i>Longitudinal Differences in Language Learning Motivation by Female Participants in English and French</i>	353
Table 8.60 <i>Descriptive Statistics for Language Learning Motivation by Gender in English</i>	355

Table 8.61 <i>Differences in Language Learning Motivation by Gender in English</i>	356
Table 8.62 <i>Statistically Significant Differences between Groups in Individual Questions in English</i>	358
Table 8.63 <i>Descriptive Statistics for Language Learning Motivation by Gender in French</i>	362
Table 8.64 <i>Differences in Language Learning Motivation by Gender in French</i>	363
Table 8.65 <i>Statistically Significant Differences between Groups in Individual Questions in French</i>	365
Table 8.66 <i>Summary of Differences in Language Learning Motivation by Gender in English and French</i>	367
Table 8.67 <i>Cross-Sectional Differences in Language Learning Motivation in English by Gender</i>	369
Table 8.68 <i>Cross-Sectional Differences in Language Learning Motivation in French by Gender</i>	370
Table 8.69 <i>Longitudinal Differences in Language Learning Motivation in English by Gender</i>	371
Table 8.70 <i>Longitudinal Differences in Language Learning Motivation in French by Gender</i>	372
Table 8.71 <i>Descriptive Statistics for English Language Level and LA by CLIL Group</i>	374
Table 8.72 <i>Normality Tests for English C-test and LAT by CLIL group</i>	374
Table 8.73 <i>Comparison of the Groups “Science” and “Economics”: Language Level and LAT</i>	375
Table 8.74 <i>Longitudinal Differences in the Groups “Science” and “Economics”: Language Level and LAT</i>	378
Table 8.75 <i>Descriptive Statistics for English Language Learning Motivation by CLIL Group</i>	380
Table 8.76 <i>Comparison of the Groups “Science” and “Economics”: English Language Motivation</i>	381
Table 8.77 <i>Longitudinal Differences between the Groups “Science” and “Economics”: English Language Motivation</i>	383

Table 8.78 <i>Most Frequent First Words for Content-relevant Prompts by Science and Economics Students</i>	385
Table 8.79 <i>Ranking of Most and Least Productive Content-Relevant Prompts by Science and Economics Students</i>	386
Table 8.80 <i>Non-shared Words in English by Science and Economics Students</i>	387
Table 8.81 <i>Frequency Distributions English by Science and Economics Students</i>	388
Table 8.82 <i>Differences in Lexical Availability in CLIL related Prompts</i>	390
Table 8.83 <i>Longitudinal Differences in Content-relevant Prompts in English and French</i>	391

List of Figures

Figure 2.1 <i>Productive version of Nation’s Vocabulary Levels Test</i>	53
Figure 3.1 <i>Gardner’s (1979) Socio-Educational Model</i>	83
Figure 3.2 <i>Gardner’s (2000) Socio-Educational Model</i>	84
Figure 3.3 <i>Gardner’s (2006) Structural Equation Representation of the Socio-Educational Model</i>	85
Figure 3.4 <i>Ushioda’s (1998) Theoretical Framework of Motivation from a Temporal Perspective</i>	95
Figure 5.1 <i>The Continuum between Content-Oriented Classes and Language-Oriented Classes</i>	187
Figure 8.1 <i>Lexical Availability in English and French</i>	255
Figure 8.2 <i>Cross-Sectional Differences in the English and French LATs</i>	257
Figure 8.3 <i>Longitudinal Differences in the English and French LATs</i>	259
Figure 8.4 <i>Motivation in English and French</i>	274
Figure 8.5 <i>Cross-Sectional Differences in Motivation</i>	277
Figure 8.6 <i>Longitudinal Differences in Motivation</i>	279
Figure 8.7 <i>Differences in LAT within Gender</i>	292
Figure 8.8 <i>Cross-Sectional Differences for Male Participants in the English and French LATs</i>	303
Figure 8.9 <i>Cross-Sectional Differences for Female Participants in the English and French LATs</i>	303
Figure 8.10 <i>Longitudinal Differences for Male Participants in the English and French LATs</i>	304
Figure 8.11 <i>Longitudinal Differences for Female Participants in the English and French LATs</i>	304
Figure 8.12 <i>Differences in LAT across Gender</i>	314
Figure 8.13 <i>Cross-Sectional Differences in the English LAT by Gender</i>	323
Figure 8.14 <i>Cross-Sectional Differences in the French LAT by Gender</i>	323
Figure 8.15 <i>Longitudinal Differences in the English LAT by Gender</i>	325
Figure 8.16 <i>Longitudinal Differences in the French LAT by Gender</i>	326
Figure 8.17 <i>Differences in Language Learning Motivation within Gender across Languages</i>	348

Figure 8.18 <i>Differences in Language Learning Motivation by Gender in English</i>	357
Figure 8.19 <i>Differences in Language Learning Motivation by Gender in French</i>	364
Figure 8.20 <i>English Language Level and LAT by CLIL group</i>	375
Figure 8.21 <i>Longitudinal Differences between the Groups “Science” and “Economics”: Language Level and LAT</i>	379
Figure 8.22 <i>English Language Motivation by CLIL Group</i>	382
Figure 8.23 <i>Longitudinal Differences between the Groups “Science” and “Economics”: English Language Motivation</i>	383

List of Abbreviations

AICLE	<i>Aprendizaje Integrado de Contenidos y Lenguas Extranjeras (CLIL)</i>
AMTB	Attitude Motivation Test Battery
AWL	Academic Word List
BEP	Bilingual Education Project
BICS	Basic Interpersonal Communicative Skills
BNC	British National Corpus
CALP	Cognitive Academic Language Proficiency
CATSS	Computer Adaptive Test of Size and Strength
CBI	Content-Based Instruction
CBLT	Content Based Language Teaching
CDS	Complex Dynamic System
CEIL	Content and English Integrated Learning
CLIL	Content and Language Integrated Learning
COCA	Corpus of Contemporary American English
DL	Disponibilidad Léxica
EFL	English as a Foreign Language
EMI	English-Medium Instruction
EMILE	<i>l'Enseignement de Matières par l'Intégration d'une Langue Étrangère (CLIL)</i>
EVST	Eurocentres Vocabulary Size Test
FI	Formal Instruction
FFL	French as a Foreign Language
FL	Foreign Language
GSL	General Service List
L1/2/3	First/Second/Third Language
L2MSS	L2 Motivational Self System

LA	Lexical Availability
LAT	Lexical Availability Task
LFP	Lexical Frequency Profile
LLOS	Language Learning Orientation Scale
LLS	Language Learning Strategies
LOPP	Lexical Oral Production Profile
LOTEs	Languages Other than English
MFQ	Motivational Factors Questionnaire
NVLT	New Vocabulary Levels Test
PALE	<i>Programa de Apoyo a La Enseñanza y el Aprendizaje de Lenguas Extranjeras</i> (Foreign Language Learning and Teaching Support Programme)
PILC	<i>Proyectos de Innovación Lingüística en Centros</i> (School Language Innovation Projects)
PPLI	Perceived Positive Language Interaction
PVLT	Productive Vocabulary Levels Test
SA	Study Abroad
SDT	Self-Determination Theory
SES	Socioeconomic Status
SLA	Second Language Acquisition
TL	Target Language
TTR	Type-Token Ratio
VLТ	Vocabulary Levels Test
VST	Vocabulary Size Test
WTC	Willingness to Communicate

Chapter 1: Introduction

Towards the end of the 20th century, Europe was acutely aware of the need to make a change to language teaching practices, which had largely proven to be ineffective at meeting the needs of a globalising world (Goris et al., 2019). It was thus decided that each citizen should be expected to be able to communicate not only in their L1, but also two other European languages (Lasagabaster & López Belouqui, 2015). In order to accomplish this, proposals were made to introduce bilingual teaching, whereby content would be taught through the medium of an FL. This approach, known as Content and Language Integrated Learning (CLIL), was seen as a way to overcome previous shortfalls in language teaching by providing a more authentic scenario for language use, giving students a reason to use the language, and economically increasing exposure to the FL without increasing demands on already busy timetables (Dalton-Puffer & Smit, 2007). This approach was greeted enthusiastically within a Spanish context, first and foremost due to desperate need to improve Spaniards' FL ability, which was lagging behind other European countries (European Commission, 2012). In addition, dual-language content teaching which was already in place in Spain's bilingual regions provided a unique starting point and experience from which to build on. The CLIL approach has consequently become widely implemented throughout Spain, with the aim of promoting multilingualism and improving foreign-language learning (Ruiz de Zarobe & Lasagabaster, 2010). However, the enthusiasm for the approach has largely outpaced research which supports its implementation (Pérez-Cañado, 2012). That is to say, in many cases, programmes have been put into place without scientific assurance that they are truly beneficial or are being implemented in an optimal way. This issue is all the more problematic in terms of the language of instruction, given that English has largely stolen the limelight, both in research and practice, leading to an evident neglect of other target languages (TLs). This has become so apparent that suggestions have been made to refer to the approach as Content and English Integrated Learning (CEIL) (Dalton-Puffer, 2011). However, given that educational centres do in fact implement CLIL programmes in other languages such as French and German, there is a clear need to carry out comparative research across additional languages, so as to determine the strengths and weaknesses of CLIL language-independently (Dalton-Puffer, Nikula & Smit, 2010).

Within this context, there are two key areas which have arisen as particularly important, namely, vocabulary and motivation. Firstly, it is well known that vocabulary is an essential component of language learning, and it has been remarked that it may well

even be the most important level of second language (L2) knowledge (Saville-Troike & Barto, 2017). As Barcroft (2004) notes, this is because communication is only possible when we use vocabulary, because students instinctively understand its importance, and because vocabulary knowledge plays a critical role in grammar development. Vocabulary also permits us to use language, while using the language allows us to increase our vocabulary knowledge (Nation, 2001). This importance evidently crosses over into the context of CLIL, where it has been suggested that CLIL teaching has a positive effect on the acquisition of content-related vocabulary (Heras & Lasagabaster, 2015). This is because students receive increased exposure to the TL and need to incorporate specific vocabulary which is related to the CLIL subject into their lexicon. However, as noted above, there is a need for research to provide further empirical evidence to support this claim. In particular, in the field of lexical availability (LA), Canga Alonso (2017) has called for a focus on LA prompts which are related to the CLIL subject, so as to determine the extent to which students are actually acquiring this vocabulary. The same researcher has also highlighted the need to include some measure of language proficiency in conjunction with LA analysis, in order to ascertain the effect of language level on the number of words that students retrieve. Secondly, motivation has long been confirmed as an essential factor in language learning, with students and teachers alike attributing language learning success or failure to whether or not the learner is motivated (Dörnyei, 2013). Research has pointed to numerous reasons why motivation plays such an integral role, such as the fact that it has a direct influence on L2 learning strategies, interaction with native speakers, language input, curriculum-related tests, general proficiency, and language retention (Oxford & Shearin, 1994). Furthermore, motivation has been found to play an even more important role in CLIL than in non-CLIL settings (Navarro & García Jiménez, 2018). This is due to the fact that it can provide a more cognitively challenging situation, promote fruitful discussion, and provide a sense of ownership in teaching and learning (Lasagabaster, 2019). However, again, it has also been remarked that this advantage is generally taken for granted, rather than supported by empirical studies (Doiz, Lasagabaster & Sierra, 2014a). Thus, in the case of both vocabulary and motivation, there is a clear assumption that CLIL offers benefits, despite the obvious need for more research to back up these claims. In relation to the CLIL/CEIL division outlined above, these issues deserve even more attention. English as a global language is increasingly viewed as a basic educational skill, crucial to professional advancement, which can result in a qualitative difference between the motivation for learning English compared with other languages (Dörnyei & Ushioda, 2013). It remains to be seen whether CLIL's purported

benefits to motivation hold true when there are multiple languages at play, and in particular when other languages are in competition with English. Recent research by Baten, van Hiel and de Cupere (2020) has also indicated that CLIL students may have larger vocabularies in English as compared to French, though there is a clear need to investigate the effect of subject matter. Bearing the above in mind, there is an evident need to explore the differences between CEIL as compared to CLIL, particularly in terms language learning motivation and its effect on vocabulary acquisition.

A further factor which is of vital importance in this context is gender. Research has by and large found that females, who generally appear to be more motivated, outperform their male peers in language learning (San Isidro, 2010). However, suggestions have been made that CLIL may lead to a potential blurring effect of gender differences (Lasagabaster, 2008), as male students may offset lower FL learning motivation with higher motivation towards the CLIL subject (Heras & Lasagabaster, 2015). As research into this levelling effect of CLIL on gender has yielded mixed findings, it is necessary to address specific motivation and interest towards the content studied in CLIL. This would allow us to verify whether there is indeed an increase in language level and motivation for different genders in this learning environment.

A final consideration that must be made in any CLIL research going forward is the potential long-term effects of this learning approach. Although a longitudinal perspective is clearly required if we are to offer a meaningful explanation of language learning development and change (Ortega & Ibarra-Shea, 2005), there has been a remarkable scarcity of longitudinal CLIL research (Pérez Cañado, 2018a; Goris et al. 2019). Calls have consequently been made to adopt such an approach so as to better understand the long-term effects of CLIL education (Ruiz de Zarobe, 2011; San Isidro, 2019).

In order to address these gaps, this thesis sets out to meet the following objectives. Firstly, it aims to determine whether there are quantitative and qualitative differences between students' LA in English and French. Secondly, it seeks to ascertain whether there are quantitative differences between students' English language learning motivation as compared with their French language learning motivation, and whether there is a relationship between this motivation and the participants' LA and language level. Thirdly, it intends to verify whether there are quantitative and qualitative differences between male and female students' LA in English and French and quantitative differences between their language learning motivation in each language. Finally, it aims to clarify whether there is an effect of CLIL instruction on the students' LA and language learning motivation in

each language. In all cases, the study seeks to analyse these factors across time in order to determine the long-term effects.

In order to meet these objectives, this study focuses on students who are enrolled simultaneously in CEIL (content through English) and CLIL (content through another language, French) in Spanish context. Within this context, it analyses the LA and language learning motivation of 9th, 10th and 11th grade students in each TL, in order to determine potential differences between English and French. This is done using three main instruments in each language: a language level C-test, a lexical availability task and a language learning motivation questionnaire. To specifically address the issues outlined above, these instruments cater to the particular learning environment of the participants in this study. This includes incorporating LA prompts that are related to the students' CLIL classes in each language and introducing questions into the motivational questionnaires that explicitly address the students' attitudes towards these classes and learning experience. Gender differences are also addressed by comparing the male and female students in each group. These various factors are further examined across the three grades by means of pre- and post-tests in order to determine whether there are changes across time. This includes a cross-sectional analysis from 9th to 10th grade and a longitudinal analysis from 10th to 11th grade.

This thesis will open with four chapters which provide the key concepts and background to this work. Chapter 2 deals with the acquisition of vocabulary in Second Language Acquisition (SLA). It begins by highlighting the importance of vocabulary and some of its main challenges. It then provides an overview of vocabulary research into the dimensions of word knowledge and reviews the main tools that have been used in measuring vocabulary. It closes with a focus on LA, providing a background and an overview of relevant research. Chapter 3 focuses language learning motivation. It first provides an introduction to motivation in SLA and a historical overview of the main theories and approaches in the field. It then provides an overview of the most pertinent L2 motivation research related to the issues of the language of study, learning context, and gender. Chapter 4 turns to the role of gender, opening with an introduction to the construct of gender in language acquisition and then specifically SLA. It then addresses prior research into gender which is related to vocabulary, motivation and CLIL. The literature review closes with Chapter 5, which focuses on CLIL. It first provides the background and key characteristics of the approach, as well as some key considerations in its adoption. It then offers a contextual background to CLIL in Spain, and specifically in the autonomous community of La Rioja. It then delves into CLIL research, highlighting

work that has been carried out in relation to vocabulary, motivation and gender. Following on from this theoretical background, Chapter 6 outlines the main rationale for carrying out this thesis, its objectives, research questions and hypotheses. The methodology and procedure that has been adopted in the study is then outlined in Chapter 7, dealing with the study design, the participants, the instruments, the data collection process, and the data analysis procedure. Chapter 8 provides an overview of the results, which are then discussed in Chapter 9. Both chapters are structured around the study's research questions, first discussing the TL influence on LA, then the TL influence on language learning motivation, next the issue of gender, and finally CLIL. The final chapter offers a summary and conclusion, highlighting the main contributions of this work and the pedagogical implications of the research findings, the limitations of the research that has been carried out, and some suggestions for further research.

Chapter 2: Vocabulary and Second Language Acquisition

This chapter will open with an introduction to the area of vocabulary, firstly outlining key reasons for its importance and secondly highlighting some of the main challenges in learning vocabulary. The third section will then provide an overview of vocabulary research into the dimensions of word knowledge, offering a detailed account of how lexical competence has been theorised in order to be analysed. The fourth section provides a review of how vocabulary has been measured, both with regards to receptive knowledge and productive knowledge. The final section will focus on an area of vocabulary which is of central importance to the thesis: lexical availability. This section will provide a background to lexical availability and its measurement, and provide a thorough overview of research that has been carried out in this area.

2.1. The Importance of Vocabulary in Second Language Acquisition

Vocabulary, or the lexicon, has been defined as “the words of a language, including single items and phrases or chunks of several words which convey a particular meaning, the way individual words do” (Lessard-Clouston, 2013, p. 3). It is just one of the six vital components of SLA, alongside morphology, phonology, syntax, nonverbal structures and discourse; however, as Saville-Troike and Barto (2017) note, it may well be the most important level of L2 knowledge for learners to develop. Barcroft (2004, p. 201) raises three points which highlight why vocabulary is so important in SLA:

1. the relationship between vocabulary and the ability to communicate
2. student perceptions about the relative importance of vocabulary
3. the critical role of vocabulary knowledge in the development of grammatical competence.

Firstly, as Wilkins (1972, pp. 111-112) astutely pointed out, “without grammar very little can be conveyed, without vocabulary nothing can be conveyed”. Essentially, a person who makes a grammatical mistake can generally still be understood, but a person who does not use a particular word or uses one with a different meaning may not be understood at all. Barcroft (2004, p. 201), for example, highlights the errors in the following sentences with the intended meaning “It snows”: *It snow. / *It nevs. Evidently, although the first sentence is ungrammatical as it lacks the appropriate agreement, the essential meaning is still transmitted. However, the second sentence clearly transmits no intelligible message.

Secondly, vocabulary may be considered significant simply because of the

importance students attribute to it. As is pointed out by Krashen (1989), learners tend to carry around dictionaries rather than grammar books, highlighting their instinctive recognition of the importance of vocabulary. According to Klapper (2008), vocabulary appears to be the aspect in which learners apply learning strategies most frequently, perhaps because they place greatest store on vocabulary. This awareness has also been shown by James (1996), who reported survey results which highlight students' particular interest in receiving vocabulary instruction.

Finally, it has also been suggested that much of what we think of as grammatical knowledge is in fact found at the level of individual words (Barcroft, 2012). For example, Healy and Sherrod (1994) found that English speakers pronounce the /ə/ before consonant sounds and /i/ before vowel sounds using information stored at the lexical level, which develops gradually as speakers are exposed to new examples of words. Serwatka and Healy (1998) showed that the speaker's ability to distinguish between count and mass nouns in English is also based on lexical knowledge. Furthermore, Barcroft (2007) carried out a study which analysed native English speakers' ability to make grammaticality judgements using real and unreal words. Results found that participants' ability to make accurate judgements decreased dramatically when working with unreal words.

In addition to these three points, Nation (2001) highlights the complementary relationship between vocabulary knowledge and language use: vocabulary knowledge enables learners to use the language, and at the same time language use enables the increase of vocabulary knowledge. In other words, learning words allows the learner to use the language and using the language allows the learner to learn words. However, despite its clear importance in language learning, learning vocabulary is by no means an easy task, as will be discussed in the next section.

2.2. Challenges in Learning Vocabulary

The task of learning vocabulary has long been acknowledged as one of the most difficult challenges of learning an FL. Sweet (1899/1964, p. 64), for example, maintained that "the real intrinsic difficulty in learning an FL lies in that of having to master its vocabulary"; a statement which still appears to hold true today. According to Laufer and Nation (2012), the reason for this difficulty may be attributed to quantitative, qualitative, and environmental or situational factors. Firstly, as opposed to grammar, which is a system of a limited number of rules, vocabulary is an open set of an infinite number of items, and thus entails a huge effort on the part of the learner to acquire this enormous

lexicon. Furthermore, it is not enough for a learner to simply memorise this sheer quantity of ever-changing and ever-growing list of words. They must also have considerable qualitative knowledge of each of the words regarding its part of speech, its derivations, inflections, collocations, and how it can be used in social situations (Milton & Donzelli, 2013). Finally, while learners receive constant grammatical reinforcement in every phrase they encounter, they may receive very little exposure to low-frequency vocabulary in particular contexts (Laufer & Nation, 2012). Thus, depending on the learning environment, learners may receive insufficient reinforcement of less common vocabulary, which requires a great deal of input in order to be acquired.

A further issue is that, despite the awareness of its importance, there has often been a distinct lack of attention to vocabulary as compared with other parts of language (Gass, 2013). Milton (2009) attributes this to three main factors:

1. a focus on grammar instead of vocabulary
2. the belief that language can be learned with a limited vocabulary
3. the consensus that vocabulary can be learned incidentally

Firstly, structuralist approaches to language teaching have traditionally sidelined vocabulary, focusing on how language rules and systems are acquired, without being overly concerned about the words to which these rules and systems apply. In essence, the volume of vocabulary is typically reduced to what is required for language presentation or to motivate the learner.

Secondly, there remains a persistent belief that proficiency can be attained even with limited lexical resources. For example, Ogden's (1930) *Basic English* suggests that just 850 words are required to learn a language. However, more recent research has shown that thousands of words are needed for comprehension. One important study by Nation (2006), for example, found that in order to achieve 98% coverage of a text, the reader would need an 8,000 to 9,000 word-family vocabulary to understand written text and a vocabulary of 6,000 to 7,000 for spoken text. In addition, regarding academic texts, a more specific vocabulary will also be required to ensure comprehension. For example, Coxhead's (1998) Academic Word List (AWL) consists of 570 word families and covers around 10% of words found in academic texts (Nation & Chung, 2011). Despite these estimations, the presumption that a language can be learned with a minimum vocabulary persists, perhaps, as Milton (2009) suggests, due to wishful thinking. In other words, given the difficulties implied in learning an FL, it is evidently more appealing to believe one must learn a thousand words as opposed to several thousand.

Finally, there appears to be a general consensus that any time spent teaching vocabulary is wasted, given that it can be learned incidentally (Milton, 2009). As is explained by Barcroft (2012), incidental vocabulary learning involves learning new words from their context without a specific intention to do so, as opposed to intentional vocabulary learning which involves actively learning new words with a specific intention to do so. For example, incidental vocabulary learning may involve picking up a new word while reading or during a conversation without specifically intending to learn it. On the other hand, intentional vocabulary learning may involve learning lists of words or completing activities with the purpose of learning L2 vocabulary. This dichotomy has also been addressed using the terms indirect and direct learning of vocabulary (Nation, 1982), implicit and explicit learning (DeCarrico, 2001), deliberate and incidental learning (Klapper, 2008), deliberate, decontextualised vocabulary learning and learning from context (Nation, 2006) and input-based learning and form-focused instruction (Laufer, 2009). Vocabulary research has largely been divided between these two sides, with some researchers seeing vocabulary learning from context as something opposed to direct vocabulary learning and teaching (Kelly, 1990). As is pointed out by Laufer (2009), this division has been one of the central concerns in vocabulary research, with researchers trying to understand how L2 vocabulary is learned, be it through L2 input, enhanced input, interaction, communicative tasks, non-communicative ‘artificial’ exercises, list learning, or repetition. On the one hand, Krashen’s (1985) Input Hypothesis provided support for indirect approaches to vocabulary learning, stating that “vocabulary in L2 was acquired subconsciously through comprehensible input, particularly through reading, when learners focused on messages, and not on individual words” (Laufer, 2009, p. 343). The argument behind this approach was based largely on the premise that L2 vocabulary development is the same as first language (L1) vocabulary development (Klapper, 2008). On the other hand, Schmidt’s (1990) Noticing Hypothesis states that simply being exposed to L2 input does not guarantee that learners will actually notice this input. Similarly, Swain’s (1985) Pushed Output, which was proposed based on immersion programme research, highlights that meaning-focused productive work in the form of sustained speaking and writing is needed to push learners to produce comprehensible output (Klapper, 2008). Such findings support the Focus on Form approach (Long, 1991; Ellis, 2001) and the argument that comprehensible input, while necessary, is not a sufficient condition for SLA (Long, 1981). Milton (2009) shares this view, highlighting the need for words to be taught explicitly, and then supplemented by meaningful

vocabulary-targeting activities outside the classroom. Other researchers, such as Nation (2001) have advocated a more balanced approach, viewing incidental and intentional learning as complementary activities, with each one enhancing the learning of the other. It has also been suggested that different approaches may be necessary at different stages of learning, with intentional vocabulary learning being seen as “necessary to reach a vocabulary size ‘threshold’ that enables incidental learning from reading” (Schmitt, 2000, p. 120). In other words, at the early stages of language learning learners may need to learn vocabulary intentionally, which will then allow them to acquire vocabulary incidentally as their level improves.

2.3. Vocabulary Research: Dimensions of Word Knowledge

Since the 1980s, there has been rapid and dynamic development in research focusing on vocabulary acquisition (Daller, Milton & Treffers-Daller, 2007). In particular, the 1990s saw a boom in research on L2 vocabulary (Laufer & Nation, 2012), with more and more researchers focusing on its investigation and acknowledging its status as a “core component of all the language skills” (Long & Richards, 2007, p. xii). Probably the most comprehensive addition to L2 vocabulary learning has been Nation’s (2001) book *Learning Vocabulary in Another Language*, which outlined four strands in balanced vocabulary learning: comprehensible meaning-focused input, form-focused instruction, meaning-focused output, and fluency development.

However, as is pointed out by Henriksen (1999), after twenty years of insights, there was a clear need for more clarity and standardisation when describing central processes in vocabulary learning and use. This need was still being discussed over a decade later by Laufer and Nation (2012), who note that a crucial factor in conducting this research and setting language instruction goals has been defining lexical competence and measuring it in a way that it is valid and reliable. The following two sections will address this issue, first introducing the notion of lexical competence, and then discussing the many distinctions that have been made in lexical knowledge.

2.3.1. Lexical Competence

Lexical competence is defined by The Council of Europe as “knowledge of, and ability to use, the vocabulary of a language” (2001, p. 110). While in theory this definition presents a very simple explanation of lexical competence, in practice its definition has proven to be much more challenging. This is perhaps, as Jiménez Catalán (2002) remarks, due to the multifaceted nature of words and the fact that a definition of lexical competence

is sustained in different dimensions, such as the linguistic, sociolinguistic, psycholinguistic and pedagogic dimensions. Lexical competence has also been discussed by different researchers under the labels of *word knowledge*, *vocabulary knowledge*, and *lexical knowledge*, adding further terminological confusion to an already complex task (Jiménez Catalán, 2002).

Jiménez Catalán (2002) justifies four main reasons why defining lexical competence is so important: it allows us to a) analyse the principles that govern vocabulary acquisition; b) discover if there is a systematic order reflected in stages of its acquisition; c) test whether these stages are different in the L1 and L2; and d) determine whether it is influenced by contextual and individual factors such as the type of input, the teaching method, age, sex, motivation and linguistic aptitude. In addition, its clarification is of significant practical importance within pedagogy, given that its understanding is vital if language teachers are to fully comprehend what is occurring in vocabulary activities and tests (Schmitt, 1995).

One of the first attempts to define what is understood by lexical competence can be found in Richards' (1976) seminal article *The Role of Vocabulary Teaching*, which outlined the following eight assumptions concerning the nature of lexical competence:

1. The native speaker of a language continues to expand his vocabulary in adulthood, whereas there is comparatively little development of syntax in adult life.
2. Knowing a word means knowing the degree of probability of encountering that word in speech or print. For many words we also know the sort of words most likely to be found associated with the word.
3. Knowing a word implies knowing the limitations imposed on the use of the word according to variations of function and situation.
4. Knowing a word means knowing the syntactic behaviour associated with the word.
5. Knowing a word entails knowledge of the underlying form of a word and the derivations that can be made from it.
6. Knowing a word entails knowledge of the network of associations between that word and other words in the language.
7. Knowing a word means knowing the semantic value of a word.
8. Knowing a word means knowing many of the different meanings associated with a word. (Richards, 1976, pp. 76-82).

These assumptions, concerned with frequency, register, syntax, derivation, association, semantic values and polysemy, served as a starting point for describing lexical knowledge in most subsequent research (Ma, 2009) and led to a trend in studying lexical competence and an effort to further define what knowing a word means (Jiménez Catalán, 2002). In doing so, researchers have opted for either a description of the various different traits comprising all aspects of word knowledge, such as that outlined above by Richards, or a global description, usually consisting of one to three dimensions (Henriksen, 1999). These various dimensions will now be dealt with the next section.

2.3.2. Dimensions in lexical knowledge

In the previous section, lexical competence was defined, quite simply, as knowing the vocabulary of a language and being able to use it. Yet despite this seemingly simple definition, the complexity of lexical knowledge has led to a wide variety of dimensions, including those outlined in Table 2.1 and discussed in turn below.

Table 2.1

Distinctions in Lexical Knowledge

Study	Dimensions		
Palmer (1921)	Productive Vocabulary	Receptive vocabulary	
Nation (2001)	Form	Meaning	Use
Bialystok & Sharwood Smith (1985)	Knowledge	Control	
Berman et al. (1968)	Potential Vocabulary	Real Vocabulary	
Laufer & Paribakht (1998)	Passive Vocabulary	Controlled-Active Vocabulary	Free-Active Vocabulary
Henriksen (1999)	Partial to Precise Knowledge Dimension	Depth of Knowledge Dimension	Receptive to Productive Dimension
Wesche & Paribakht (1996)	Breadth	Depth	
Read (2004)	Precision of Meaning	Comprehensive Word Knowledge	Network Knowledge

Firstly, a distinction is generally made between productive, or active, knowledge, and receptive, or passive, knowledge. As Milton (2009) points out, this is a division of knowledge which has been discussed as far back as the 1920s by Palmer (1921), although there is still an absence of generally accepted definitions of productive and receptive knowledge (Pignot Shahov, 2012). However, according to Nation (2001, pp. 24-25), “receptive vocabulary use involves perceiving the form of a word while listening or reading and retrieving its meaning”, whereas “productive vocabulary use involves wanting to express a meaning through speaking or writing and retrieving and producing the appropriate spoken or written word form”. Thus, while receptive knowledge is essentially concerned with understanding a word, productive knowledge is concerned with producing it. More specifically, according to Gass (2013), the two types of knowledge include the following differences outlined in Table 2.2.

Table 2.2

Differences between Receptive and Productive Knowledge

Receptive Knowledge	Productive Knowledge
<ul style="list-style-type: none"> • Recognition of the word • Knowing the general meaning • Knowing the specific meaning in context • Knowing its component parts • Knowing its connotations • Knowing what it generally occurs with • Knowing the opposite 	<ul style="list-style-type: none"> • Knowing how to pronounce and spell it correctly • Knowing the precise meaning in a variety of contexts • Knowing specific collocations and what other words can and cannot be used alongside it • Knowing the precise context of use

Note. Adapted from *Second language acquisition: An introductory course* (4th ed. p. 197), by S. M. Gass. 2013. Routledge. Copyright 2013 by Routledge.

According to Laufer and Goldstein (2004, p. 405), there is “no consensus as to whether this distinction is dichotomous or whether it constitutes a continuum”. Some researchers, such as Melka Teichroew (1982) and Wesche and Paribakht (1996), believe that these types of knowledge should be viewed on a continuum, beginning with a superficial familiarity with a word and later being able to use the word correctly in free production. Others, however, such as Meara (1997) believe that these types of knowledge represent

different types of associational knowledge and consequently cannot be a continuum. Regardless of which stance is taken, be it a bi- or multi-dimensional representation or viewing receptive and productive knowledge as a continuum or as two separate poles, they all, at the very least, accept the receptive and productive dimension (Pignot-Shahov, 2012).

A second important distinction is proposed by Nation (2001), who refers to inter-related *sub-knowledges*, that is to say, knowledge regarding the word's form, meaning and use. As outlined in Table 2.3, these three categories are further divided into three subcategories, each of which involves both a productive and receptive element.

Table 2.3

What is Involved in Knowing a Word

Form	Spoken	R	What does the word sound like?
		P	How is the word pronounced?
	Written	R	What does the word look like?
		P	How is the word written and spelled?
	Word parts	R	What parts are recognisable in this word?
		P	What word parts are needed to express the meaning?
Meaning	Form and meaning	R	What meaning does this word form signal?
		P	What word form can be used to express this meaning?
	Concept and referents	R	What is included in the concept?
		P	What items can the concept refer to?
	Associations	R	What other words does this make us think of?
		P	What other words could we use instead of this one?
Use	Grammatical functions	R	In what patterns does the word occur?
		P	In what patterns must we use this word?
	Collocations	R	What words or types of words occur with this one?
		P	What words or types of words must we use with this one?
	Constraints on use (register and fluency)	R	Where, when, and how often would we expect to meet this word?
		P	Where, when, and how often can we use this word?

Note. In column 3, R = receptive knowledge, P = productive knowledge. From *Learning vocabulary in another language* (p. 27), by I. S. P. Nation, 2001, Cambridge University Press. Copyright 2001 by Cambridge University Press.

Thirdly, Bialystok and Sharwood Smith (1985) drew a distinction between knowledge and control. While knowledge was defined as “the way in which the language system is represented in the mind of the learner”, control was seen to be “the processing system for controlling that system during actual performance” (Bialystok & Sharwood Smith, 1985, p. 104). The authors used the analogy of a library to explain the difference: knowledge is the books and the way in which they are organised, whereas control is how the books are accessed. While such an analogy is useful in its application to the lexicon, Gass (2013) also points out that it is particularly problematic in that it does not take into account the dynamic and changing nature of the lexicon.

Fourthly, Berman, Buchbinder and Beznedezych (1968, as cited in Palmberg, 1987, p. 21) distinguish between potential vocabulary and real vocabulary. The former addresses vocabulary which is easily recognizable to the learner, despite the fact that it is the first time it has been seen in the L2. For example, many technical or scientific terms are very similar across languages, thus, the learner will evidently recognise such terms even though they have not yet been learnt explicitly. The latter deals with vocabulary which, after and due to exposure, the learner has become familiar with.

A fifth distinction is that of passive, controlled-active and free-active vocabulary (Laufer & Paribakht, 1998). The basic passive or receptive knowledge refers to the understanding of the most frequent and core meaning of a word. The authors give the example of the word “solution”, with its most common meaning of “solution of a problem” rather than its more technical one of a “chemical solution”. They then distinguish between two types of productive knowledge: controlled and free. Controlled productive knowledge involves cued recall, producing words when prompted by a task, for example, completing the word “fragrant” in the sentence “the garden was full of fra___ flowers” (Laufer, 1998, p. 257). Free productive knowledge does not involve any specific prompts, but instead focuses on the spontaneous use of words, such as in the case of free composition. Laufer and Paribakht (1998) highlight the usefulness of distinguishing between free-active vocabulary (words which are chosen voluntarily) and controlled-active vocabulary (words which can be used if required), given that learners may avoid using particular words in free expression when they find them problematic or feel uncertain about using them.

Henriksen (1999, p. 304) proposed three dimensions of lexical competence which are “separate but related”, namely the partial to precise knowledge dimension, the depth of knowledge dimension and the receptive to productive dimension. As is explained by

Leńko-Szymańska (2020), the first refers to the learner's degree of familiarity of the word, the second to how well the word is embedded in the learner's network of sense relations with other words, and the third to the level of access to a word or the ability to use it.

Lexical competence may also be analysed in terms of breadth or size, that is, its quantity- the number of words known, and depth, that is, its quality- how well it is known (Wesche & Paribakht, 1996). Although for a long time research focused solely on the size of vocabulary, further development in vocabulary acquisition research has come to see vocabulary size as only one aspect of this knowledge and has realised the importance of depth in understanding vocabulary acquisition (Yue & Fan, 2016). This distinction between quantity and quality has since been widely used to investigate lexical knowledge, particularly with regards to reading comprehension (Gass, 2013). However, there is still a great deal of debate regarding the relationship of the two constructs. Some have argued that the two facets are intrinsically related, given that as learners expand the total number of words that they understand, they at the same time learn more about those words which they encounter (Read, 2004). Vermeer (2001) has even argued that there is essentially no difference between the two, given that it is through knowledge of related words that learners are able to understand the meaning of individual words. On the other hand, using regression analyses, Qian (1999, 2000) has shown that depth has a unique explanatory power compared to breadth alone. As Schmitt (2014) highlights, the relationship between these two contrasts will ultimately depend on how they are conceptualised and measured.

With regard to depth, Read (2004) further argues that this one single term is insufficient, and that it should instead be replaced by more specific definitions which better reflect what is actually being assessed by particular vocabulary instruments. Specifically, he proposes three distinct lines of development in the application of depth to L2 vocabulary acquisition: precision of meaning, comprehensive word knowledge, and network knowledge. Precision of meaning refers to how well the meaning of the word is known, that is, having a vague idea as compared with a more elaborated and specific knowledge. Comprehensive word knowledge refers to knowing aspects of the form, meaning and use of the word, that is, its orthographic, phonological, morphological, syntactic, collocational and pragmatic characteristics. Finally, network knowledge refers to the incorporation of a word into a lexical network and fluency of access, that is, being able to link it to and differentiate it from other related words.

Such variety and lack of agreement in distinctions and definitions in vocabulary knowledge show a clear reflection of the complexity of what at first appears to be a

somewhat simple issue. In addition to understanding these distinctions, when assessing vocabulary, it is also evidently important to first establish the purpose of the assessment, so as to understand what it is we are measuring and consequently determine the most appropriate instruments to use. Before addressing the measurement of vocabulary knowledge in the next section, this section will conclude by outlining Read's (2000) three dimensions of vocabulary assessment (Table 2.4), which can be used to better understand the variety of assessment procedures currently in use and thus enable us to cater these tests more specifically to their intended purpose.

Table 2.4

Dimensions of Vocabulary Assessment

Dimensions of Vocabulary Assessment		
Construct	Discrete	Embedded
Range	Selective	Comprehensive
Context	Context-Independent	Context-dependent

Note. Adapted from *Assessing Vocabulary* (p. 9), by J. Read, 2000, Cambridge University Press. Copyright 2000 by Cambridge University Press.

The first dimension concerns construct, and distinguishes between discrete tests (those which measure vocabulary as an independent construct) and embedded tests (those which measure vocabulary as part of some other, larger construct). For example, a multiple-choice vocabulary test aims specifically to measure vocabulary. On the other hand, a text in which students must write a proposal will assess several different areas, just one of them being vocabulary, to measure the students' writing ability. The second dimension deals with the range of vocabulary in the assessment; be it selective (whereby specific items are the focus) or comprehensive (whereby all lexical content of a test is taken into account). Selective tests include those which seek to assess specific target words, such as the vocabulary the students have covered in a unit. An example of a comprehensive test is a speaking exam, where the examiners are not listening for particular words, but instead judge the quality of the students' overall vocabulary use. The third dimension relates to context, distinguishing between context-independent (expected response can be produced without referring to any context) and context-dependent tests (contextual information must be taken into account in order to produce the expected response). The former can be seen in tests where the student is given a word in isolation and must supply its meaning without any contextual cues, whereas the latter asks for the meaning of a word after providing some kind contextual information; for example, asking for the meaning of a

word which has appeared in a reading comprehension.

As is pointed out by Read (2000, p. 13), rather than intending to form a comprehensive model of vocabulary assessment, these three dimensions aim to offer a basis for “locating the variety of assessment procedures currently in use within a common framework”. In addition, they provide a bridge between tests which treat words as discrete units and those which assess vocabulary in a more integrative task-based testing context.

Bearing in mind these dimensions, the following two sections will outline some of the main measures that have been used to investigate vocabulary knowledge.

2.4. Vocabulary Research: Measurement

This section outlines the main methods that have been used in vocabulary research in the measurement of receptive and productive knowledge. It will open by outlining and explaining some key terms and issues which will be necessary in the following sections, including types, tokens, lemmas, word families, and the use of word frequency lists.

Nation (2001) outlines four different ways of deciding how words should be counted: tokens, types, lemmas and word families. Firstly, tokens are the number of different words in a text while types are the number of different words. For example, the sentence “The cat sat on the mat” contains six tokens and five types, given that the word *the* is repeated twice (Milton, 2009). Next, a lemma is “a set of lexical forms having the same stem and belonging to the same major word class, differing only in inflection and/or spelling” (Francis & Kučera, 1982, p. 1). In English, these inflections are made up of the plural, third person singular, present tense, past tense, past participle, -ing, comparative, superlative and possessive (Bauer & Nation, 1993). For example, the base form *give* also has an associated set of inflexions *gives*, *giving*, *gave*, and *given*. A similar notion is the word family which, according to Bauer and Nation (1993, p. 253), “consists of a base word and all its derived and inflected forms that can be understood by a learner without having to learn each form separately”. A key difference between the lemma and the word family is the inclusion of derivations in the latter: unlike lemmas, a word family will also include affixes such as *-ly*, *-ness* and *un-* (Nation, 2001). For example, with regard to the word *like*, the lemma includes the inflected forms *like*, *likes* and *liked* whereas the word family will contain the inflected forms *like*, *likes*, *liked* as well as the derivations *likely*, *likeness* and *unlike*. However, although similar, Brown (2013) argues that a word family is not simply a lemma with derivations, but rather is a fundamentally different unit. An additional difference between the two is outlined by Nation and Meara (2013): while a

word family contains words with different parts of speech, a lemma does not. Thus, *approach*, *approaches*, *approached* and *approaching* are all members of the same lemma: they have the same stem, include only the stem and inflected forms, and are all verbs. However, the noun forms *approach* and *approaches* are classified as a different lemma given that they are a different part of speech.

When it comes to deciding how to count (word type, lemma, word family and level of inclusiveness), Nation (2016) highlights the importance of determining the purpose of the counting and the intended users of the information. As is suggested by Nation and Meara (2013), while the word family is the best unit when counting learners' receptive knowledge, the word type, or possibly the lemma, is the best unit for counting productive knowledge. Furthermore, with regard to counting using word families, Nation (2001) highlights one major issue, namely, how to determine what should be included in a word family and what should not. Bauer and Nation (1993) maintain that an important principle concerning word families is that, given that the base word is known, other members of the family will be easily recognised. However, the knowledge that learners have of prefixes and suffixes will evidently vary according to their level, with elementary and transparent words being available to the learner at a lower level and less obvious possibilities becoming available as their level improves (Nation, 2001). In other words, what is considered a member of a word family will increase as proficiency develops. With this in mind, using the 1,000,000 token Lancaster-Oslo-Bergen corpus, Bauer and Nation (1993, p. 255) proposed a series of seven levels which could be used "to provide a consistent description of what should be considered to be part of a word family for readers at different levels of morphological awareness". Brown (2018) provides a very useful summary of these levels, as outlined in Table 2.5 below. As Bauer and Nation (1993) note, this scheme is designed for practical purposes with a focus on recognising written words. They admit that the divisions between the levels are arbitrary, given that certain affixes may appear in more than one level, but that they should represent possible "steps along a cline" in recognising the various words in a family (Bauer & Nation, 1993, p. 257). Thus, words at Level 2, such as *wastes*, are expected to be relatively simpler to recognise than words at Level 6, such as *wastage*. As is pointed out by Gardner (2007), one clear advantage of the scheme is its usefulness in analysing learners' ability to associate morphologically connected words in a consistent manner. However, she also highlights some problems, including the unclear distinctions between levels, lack of consideration for the relative difficulty of affixes as compared to suffixes, inadequately

addressing the role of stems in derived forms, and the linear assumption of acquisition of morphologically-related words.

Table 2.5

Summary of Bauer and Nation's (1993) Word Families Scheme

Level	Description	Number of affixes	Examples of affixes	Examples of forms
1	Each form is a different word			
2	Inflectional suffixes.	8	<i>-ed, -s</i>	<i>heat-ed</i> <i>waste-s</i>
3	The most frequent and regular derivational affixes.	10	<i>able, -er, non-</i> <i>(each with restricted uses)</i>	<i>heat-er</i>
4	Frequent, orthographically regular affixes.	11	<i>-al, -ful, in-</i> <i>(each with restricted uses)</i>	<i>waste-ful</i>
5	Regular but infrequent affixes.	50	<i>-age, -ally, ante-</i>	<i>wast-age</i>
6	Frequent but irregular affixes.	12	<i>-ee, -ic, pre-</i>	<i>pre-heat</i>
7	Classical roots and affixes.			

Note. From “Examining the Word Family through Word Lists”, by D. Brown, 2018. *Vocabulary Learning and Instruction*, 7 (1), pp. 14–34. Copyright 2018 by Vocabulary Learning and Instruction.

A final important component in many vocabulary tests is the use of word frequency lists. These lists, which are built from corpora, “rank words based on how often they appear in writing or speaking” and “give teachers and learners an idea of the most useful vocabulary to study” (Spiri, 2008, p. 22). The utility of such lists is evidently dependent on the quality of the lists themselves. Nation and Waring (1997), for example, highlight six factors which should be considered when developing a resource list of high frequency words: a) representativeness (lists should be based on both written and spoken corpora); b) frequency and range (a word should not only be frequent, but also occur frequently across a wide range of texts); c) word families (criteria regarding how members of the same family are counted should take into account the purpose of the list); d) idioms

and set expressions (items larger than words should likely be included if they are sufficiently frequent); e) range of information (information on the word such as form, meanings, collocations, etc., should be included); and f) other criteria (in addition to frequency and range, other factors such as ease or difficulty of learning, necessity and cover should also be addressed).

One of the most influential word frequency lists has been West's (1953) General Service List (GSL) (Burkett, 2015), which contains 2000 headwords based on the frequency figures of a 5,000,000 word written corpus. As is highlighted by Moreno Espinosa (2003), the GSL was the predominate word frequency list for around 50 years and laid the groundwork for computer programs such as VocabProfile (Cobb, 2018) which carry out lexical analysis. More recently, Nation (2012) developed the British National Corpus (BNC)/Corpus of Contemporary American English (COCA) word family lists, which consist of 29 word family lists representing both British and North American English and which are partly based on spoken corpora. Twenty-five lists contain word families based on frequency and range data, and four additional lists are made up of proper nouns, marginal words, transparent nouns and acronyms. As is remarked by Webb and Sasao (2013, p. 267), as compared with older lists, "the new BNC/COCA lists should be representative of current English and provide a far better indication of the vocabulary being used by native speakers today". VocabProfile also provides a 25 k-list in French: the French v.5 corpora. The lists are the result of work carried out by Lonsdale and Le Bras (2009) in the creation of their corpus-frequency based *Frequency Dictionary of French: core vocabulary for learners*. The French v.5 includes a balance of both written and spoken material and employs criteria of both frequency and range. However, unlike the BNC/COCA lists, it uses the lemma as the highest level unit rather than the family. A final important list is Coxhead's (2000) aforementioned AWL, a list of 570 word families representing high-frequency academic vocabulary from a range of disciplines including Humanities, Science, Commerce and Law (Nation, 2004).

Bearing in mind the above, the following two sections will now outline the main methods that have been used in vocabulary research in the measurement of receptive knowledge and productive knowledge.

2.4.1. Measuring Receptive Knowledge

As is pointed out by Yue and Fan (2016), understanding how to measure word knowledge is critical if we are to produce reliable and valid research; yet to date there has been a lack of consistency in assessing this knowledge. In the early days of lexical measurement in the 1980s, the aim of most studies was to determine the vocabulary size of native speakers (Laufer & Nation, 2012). However, due to poor methodological design, such as not having a representative sample of words to test, most greatly overestimated the participants vocabulary size, with some, for example Seashore and Eckerson (1940), claiming that the university students in their studies had vocabulary sizes of up to 200,000 words.

Having addressed these methodological issues, more recent studies from the 1990s have suggested more realistic figures, finding that native speakers acquire around 1000 word families per year, up to the age of around 20 (Nation, 2006). The 1990s also saw the first systematically designed tests for measuring the vocabulary size of non-native speakers (Laufer & Nation, 2012). Harrington (2018) outlines two approaches to measurement of vocabulary size: the Vocabulary Levels Test (VLT)/Vocabulary Size Test (VST) approach and the Yes/No Test approach. While both approaches aim to measure vocabulary size using frequency bands, the VLT/VST approach is characterised by a multiple-choice format in which a word is presented with alternative definitions to choose from, whereas the Yes/No Test approach has the learner simply indicate whether they know the word or not. A third approach, which although it is not included by Harrington (2018) is worthy of mention, is that which uses L2-L1 translation tests. Each approach will be now discussed in turn, outlining the main passive vocabulary tests that fall under each one.

2.4.1.1. The VLT/VST Approach. Nation's VLT (1983, 1990), revised by Schmitt, Schmitt and Clapham (2001), measures written receptive vocabulary knowledge using words at the 2,000, 3,000, 5,000, and 10,000 frequency levels and an academic vocabulary level, which in the revised version was based on Coxhead's (2000) AWL. In the revised VLT, each section includes 30 items in a multiple-matching format, with three items representing 100 words of any particular frequency band (Kremmel & Schmitt, 2018). It presents learners with six words in a column on the left and the corresponding meaning senses of three of these words in a column on the right. Learners must match each of the three words on the right with a single word from the left-hand column. Amidst

a lack of reliable tests for measuring vocabulary size, Meara (1996) called the VLT the closest thing to a standard test available in vocabulary assessment. However, one drawback of the VLT is the fact that it samples words from only four frequency bands and from there makes estimates concerning the 1K, 6K–9K and higher bands. In order to overcome this issue, the VST (Nation & Beglar, 2007) was introduced in order to provide a better estimate of overall vocabulary size (Nation, 2012). The VST is a four-alternative multiple-choice test in which learners are shown target words which appear in non-defining contexts. Distracters include similar words to the word being tested or words which are of higher frequency. Unlike the VLT, it allows each item to be tested independently, and also expands the range of frequency bands tested by including ten items from the 1K up to the 13K or 14K band if the BNC.

In a further response to some of the earlier drawbacks of the VLT, including its potentially outdated frequency lists and that it does not measure knowledge of the most frequent 1000 word families (Webb & Sasao, 2013), McLean and Kramer (2015) recently devised the New Vocabulary Levels Test (NVLT), a revised version of the VLT which also tests the first 1,000-word frequency level, uses the newer and larger BNC/COCA word lists, and makes improvements to the item format. A potential Japanese bilingual version of the NVLT has also been proposed (McLean & Kramer, 2016), and McLean, Kramer and Beglar (2015) have also put forward the Listening Vocabulary Levels Test (LVLT), a phonological version of the VLT which measures knowledge of words at the 1K–5K bands and the AWL.

One advantage of the VLT/VST approach is that it goes a step beyond the passive recognition that will be seen in the Yes/No approach, having the participant match words with the correct meaning. The tests have also been widely used and their validity and reliability has been examined in numerous studies (e.g., Read, 1988; Beglar & Hunt, 1999; Schmitt et al., 2001; Beglar, 2010; Culligan, 2015) and continue to be improved and adapted, as can be seen in the implementation of newer versions. However, it has also been pointed out that tests like the VLT include a large degree of guesswork and calculation which may not be reflected in the results, as shown quite clearly in Kamimoto's (2005) speak aloud protocols with learners taking this test. Thus, different guessing strategies may consequently produce considerable differences in learners' scores, which evidently need to be taken into consideration. The VLT is also relatively more time-consuming, taking up to six times as long as the Yes/No approach (Harrington, 2018).

2.4.1.2. The Yes/No Test Approach. One of the earliest receptive vocabulary tests in which learners had to state whether or not they knew a word was Dale's (1965) Four Stage Vocabulary Strength Test. Using a 4-point self-report scale, learners indicated how well they know items of vocabulary:

Stage 1: I never saw it before.

Stage 2: I've heard of it, but I don't know what it means.

Stage 3: I recognise it in context: it has something to do with...

Stage 4: I know it.

One evident issue with the test is that when the learner selects Stage 4, there is little information as to the extent which the word is known. Wesche and Paribakht's (1996) Vocabulary Knowledge Scale (VKS) offers an improvement of Dale's test, using a five-level scale which has the learner provide evidence for knowing a word by either supplying a synonym or translation, or using the word in a sentence. Thus, in a situation where a participant is unsure, they can state this directly, rather than simply choosing between yes and no.

Meara and Buxton (1987) also introduced a test with a similar approach, called the Yes/No Test, as an alternative to the multiple-choice tests discussed in the previous section. The test presents a list of target words and test-takers are instructed to cross out the words which they do not know well enough to say what they mean. While no attempt was made to determine if the test-taker did indeed know the word which they claimed to, the test controlled for guessing by including pseudo-words which conformed to the orthographic and phonological rules of the language, such as the French non-words *fombe*, *étoulage*, and *ponte* (Harrington, 2018). Meara and Jones (1990) later developed a computerized version of the test, called the Eurocentres Vocabulary Size Test (EVST). Based on a yes/no format, the test has learners look at an isolated word and decide if they can provide a meaning for that word, without doing so. Learners are tested on around 150 words, so as to form an estimate of their knowledge of the most frequent 10,000 words.

Meara and Milton's (2003) X-Lex test was introduced to assess knowledge of words at the 1K–5K range, using frequency bands from work by Hindmarsh (1980) and Nation (1984). Similar to the EVST, learners are shown 120 words: 20 real words from each frequency band and 20 which are invented but are made to look like real words. The learners must tick the words that they know or can use. There is also an advanced version of the X-Lex, called the Y-Lex (Meara & Miralpeix, 2006), which assesses knowledge of words at the 6K–10K range. Finally, Milton and Hopkins (2005) have devised a parallel

phonological form of the X-Lex test, the AuralLex test. It is identical to the X-Lex, except for the fact that instead of seeing the words, the learner hears them. Words can be heard as many times as the learner needs. More recently, Meara and Miralpeix (2015) have introduced V_YesNo, a free self-contained program based on the EVST, available from the *_lognostics* website. While it uses the same test items and approach to correction for guessing, it has a shorter and simpler structure than the EVST. In addition, by increasing the number of pseudowords, so that real and imaginary words occur in equal numbers, the test also reduces the amount of strategic guessing by participants (Meara & Miralpeix, 2016).

Milton (2009) highlights a number of advantages in tests like the EVST and X-Lex. Firstly, they are relatively quick and easy to construct, provided there is a suitable frequency list available. As a result, it is also possible to test a large number of words and have a larger sample size, making the results more reliable. The tests can also be relatively brief, which avoids participants getting bored, and can test words from a wide range of frequency bands. In addition, tests with both orthographic and aural versions, such as X-Lex and AuralLex can allow for the comparison of both the spoken and written form of receptive vocabulary knowledge. For example, Milton and Hopkins (2006) used X-Lex and AuralLex to compare the phonological and orthographic English vocabulary sizes of Greek and Arabic native speakers. Results showed that each aspect developed differently with language level and that orthographic word recognition was much greater than phonological word recognition in very proficient learners. On the other hand, an obvious disadvantage of receptive tests such as the EVST is the degree to which learners can guess and how to deal with words which learners are unsure of. Evidently, the validity of the tests relies to a large degree on how the participants interact with them, and thus they may not work equally well with all learners.

Regarding the comparability of the VLT/VST and Yes/No approaches, Mochida and Harington (2006) examined the performance of university-level English L2 participants on the same test items using both the Yes/No format and the VLT format. Despite the differences in item presentation and response type, participants' performance was very similar across all frequency levels, indicating that both approaches yield similar vocabulary size measures.

2.4.1.3. The Translation Approach. A final approach to measuring vocabulary size is using L2-L1 translation, whereby participants are given a list of words in their TL and

have to supply the translation in their L1. For example, Webb (2009) used L2-L1 translation in investigating the receptive vocabulary knowledge of Japanese students of English. Participants were presented with 90 items in English and had to provide the Japanese equivalent on a blank line next to each word. This study also investigated productive vocabulary using translation methods so as to be able to compare the two, as will be discussed in the next section.

Other researchers have also created translation tests which provide more context, such as the English to Indonesian translation task outlined by Nurweni and Read (1999). In this task, participants were presented with 200 items, and had to supply an Indonesian equivalent for the target words in italics, as in the following examples:

He was born in *February*.

That was the *last* event of the day.

Laufer and Goldstein's (2004) Computer Adaptive Test of Size and Strength (CATSS) is also of particular interest as it uses translation tasks to investigate both passive recall and passive recognition, as well as productive recall and recognition, which again will be discussed in the next section. The passive recall test consists of a list of words in the participants' L2, which they must translate into their L1. For some translations, the first letter is provided. The passive recognition is a four-alternative multiple-choice test in which the participant is provided with a word in their L2 and must circle the correct translation in their L1. Using the same lists and frequency bands as Smith et al.'s (2001) VLT, the CATSS includes five levels of vocabulary and 30 items at each level for a total of 150 words. As each word is tested in four modalities, there are a total of 600 items on the entire test. As was mentioned above with regard to the VLT, there has been a more recent realisation that the older frequency lists that were used may no longer be appropriate, prompting the new CATSS (Aviad-Levitzky, Laufer & Goldstein, 2019) which, as in the case of the NVLT, uses the newer BNC/COCA lists.

One advantage of such translation tasks is that the test designer can control, to some extent, the answer that will be produced, which also leaves little room for subjectivity in marking (Milton, 2009). However, Stubbe (2015) points out that although individual word translation tests are considered a reliable method of determining students' lexical knowledge, they are a somewhat more laborious task in terms of marking than the Yes/No Test. Thus, from the point of view of classroom teachers, the Yes/No approach

may be more appealing given that it requires less time to mark.

Using these three approaches, the array of tests for measuring passive vocabulary knowledge outlined above has produced credible estimates of learners' knowledge and progress with regards to vocabulary breadth (Milton, 2009). Pignot-Shahov (2012) also highlights the advantage of these receptive tests, given that they allow us to predict the amount of vocabulary which is learned throughout language courses and compare the number of words learned by students at the same level but in different countries. Following on from these receptive tests, the following section will now turn to the measurement of productive knowledge.

2.4.2. Measuring Productive Knowledge

It is now generally understood that the number of words that L2 learners know in their TL is higher than the number of words they can actually use, that is, their receptive vocabulary is larger than their productive vocabulary (Fan, 2000). Concerning the actual extent of this gap, Melka (1997) highlights some differences between the gap in the L1 as compared with the L2. For English L1 speakers, receptive vocabulary has been found to be either five times that of productive vocabulary for the average speaker (Chamberlain, 1965) or four times that of productive vocabulary in Year Two university students (Hartmann, 1946). On the other hand, for L2 learners this gap appears to be relatively smaller, with receptive vocabulary being estimated as double, or more, than that of productive vocabulary (Eringa, 1974). Fan (2000) also highlights some discrepancies in other findings: while work carried out by Laufer (1998) indicated a ratio of 89% in 10th graders and 73% with 11th graders, indicating that the gap between the two increases as learners progress, results by Morgan and Oberdeck (1930) suggested the opposite, finding a narrower gap at the end of the experimental stage. In Fan's (2000) own study, it was found that while one group of participants could recall up to 80% of the words they recognised, another could only recall around 50%, thus bringing into question the generalisation that "L2 students can recall X% of the words they recognise". A more recent study by Webb (2008) again found receptive vocabulary size to be larger than productive vocabulary. The findings also suggested that receptive vocabulary may be indicative of productive vocabulary, in that learners with a larger receptive vocabulary will likely know more of those words productively than learners whose receptive vocabulary is smaller.

The idea that productive vocabulary is smaller intuitively seems logical, firstly,

because more word knowledge is needed and, secondly, due to the contextual nature of these components which take a long time to develop (Schmitt, 2014). It has also been pointed out that when it comes to the measurement of productive vocabulary, it is relatively more difficult to characterise and operationalise this quality of knowledge in order to measure it successfully (Milton, 2009). This has led to a wide variety of approaches to measuring productive vocabulary knowledge, including translation and elicitation methods, statistical analysis of free production in speech or writing, association tests, and measures of automaticity (Milton, 2009), each of which will now be addressed in turn.

2.4.2.1. Translation and Elicitation. Translation and elicitation methods offer an example of Laufer and Paribakht's (1998) controlled productive knowledge, and include tasks such as translating lists of words, C-tests and gap-fill exercises such as in Laufer and Nation's (1995) productive vocabulary test.

Measuring productive knowledge using translation extends far back into the last century (Milton, 2009). The procedure is practically the same as that described in the receptive translation tests, but rather than translating from the L2 to the L1, the participant translates from the L1 to the L2. Thus, a list of words in the participant's native language is given to the learner who has to provide an FL equivalent. For example, in Webb's (2008) study mentioned above, Japanese natives were provided with the Japanese equivalent of words such as "bubble", "gasoline" or "bruise", and had to provide the English translation on a blank line which appeared next to the L1 cue. Quite often, productive translation tests have been carried out alongside receptive tests in order to determine the differences between the size of learners' productive and receptive vocabulary (Henriksen, 1999). For example, Webb's (2008) study mentioned above which used both L1 to L2 and L2 to L1 translation tasks found receptive vocabulary to be larger than productive vocabulary. As outlined in the previous section, Laufer and Golstein's (2004) CATSS uses both receptive and productive translation tasks, testing both recall and recognition so as to be able to compare these different aspects of vocabulary knowledge. As in the passive components of the test, participants are tested on productive recall, whereby they are given a list of words in their L1 which they must translate into the L2, and productive recognition, whereby an item is provided in their L1 and they choose the correct translation in their L2 from four options. Results of the study found that growth in vocabulary knowledge was different for different strength modalities,

that the ability to recognise words preceded the ability to recall them, and that passive recall of meaning was the best predictor of class success. Finally, translation tests have also been used to determine the reliability of recognition tests. For example, Eyckmans, van de Velde, van Hout and Boers (2007) used a receptive Yes/No vocabulary size test alongside a productive translation test and found that the participants were able to translate only 50% of the words which they had claimed to recognise.

The C-test is a text completion test based on the reduced redundancy principle, meaning that “learners’ proficiency can be measured via the rate of successful restorations of the missing message elements” (Grujić & Danilović, 2012, p. 2). It consists of a series of short authentic texts in which the second half of every second word is removed. Their development came about in order to deal with several issues with the cloze test, including length, content specificity, validity, reliability, bias and subjectivity in scoring, and assessment of difficulty (Alderson, 1979/1983; Klein-Braley, 1981). Since their development in the 1980s (Grujić & Danilović, 2012), C-tests have been used extensively as a means of testing language proficiency (Daller, van Hout & Daller-Treffers, 2003) in various languages and an array of different language learning contexts (McManus, 2011). In general, these tests are made up of between four to six texts, each containing twenty to twenty-five gaps, and participants are usually given a maximum of five minutes per text. Their popularity can largely be attributed to their straight-forward administration, scoring, high reliability and interpretation of results (Raatz & Klein-Braley, 2002; Eckes & Grotjahn, 2006). However, these numerous advantages have also come alongside a large degree of scepticism concerning their validity and criticism regarding the exact nature of what they are measuring. In particular, researchers have questioned whether C-tests may simply be a measure of reading ability (Cohen, Segal & Weiss Bar-Siman-Tov, 1985) and micro-level skills (Stemmer, 1991). Extensive research has since been carried out to address these concerns, focusing mainly on psychometric qualities such as reliability and validity given their importance in measuring devices (Farhady & Jamali, 2006). As is pointed out by McManus (2011), these studies have proven time and again the ability of the C-test to tap macro-level skills and processing (see, e.g., Grotjahn & Stemmer, 2002; Singleton & Singleton, 2002; Eckes & Grotjahn, 2006; Kontra & Komos, 2006; Khoshdel-Niyat, 2017), leading Hastings (2002, p. 24) to conclude that “the value of C-testing as a measure of global proficiency in a L2 has been demonstrated too many times to be open to dispute”. Most recently, Norris (2018) has provided detailed accounts of the development of C-tests for providing efficient measures of FL proficiency in eight

different languages. The final, five-test C-tests for all languages have again demonstrated not just impressive psychometric qualities, but also strong relationships with criterion variables such as learner self-assessments and instructional levels (Norris, 2018).

Finally, in Laufer and Nation's (1995, 1999) Productive Vocabulary Levels Test (PVLTL), learners' productive vocabulary is tested by having them complete an unfinished word in a number of different, unrelated sentences. As can be seen in the following example provided by Waring (1997) in Figure 2.1, the learner is presented with a series of sentences, each with a missing word, and is given the first two letters of the words to precipitate the exact word which is being tested.

Figure 2.1

Productive version of Nation's Vocabulary Levels Test

They always seem to ag_____ about what to do at the weekend.

Scientists are worried about the amount of co_____ in our food nowadays.

He's not a very bright child, he's about av_____.

Note. From "A comparison of the receptive and productive vocabulary sizes of some second language learners" by R. Waring, 1997, *Immaculata*, 1(1), 53-68. Copyright 1997 by Immaculata.

Translation and elicitation tests evidently have the advantage of being relatively easy both to construct and mark and, as mentioned above, are also well suited for comparison with receptive vocabulary tests. However, they have also come under criticism. For example, one disadvantage of translation tests is that they may be somewhat artificial, given the absence of the reality of communicative language. It has also been suggested that, despite aiming to measure the same type of knowledge, there is some degree of disparity across productive vocabulary tests. For example, Milton (2009) highlights a model created by Fitzpatrick (2007), which outlines the activation of the word in three different productive tests (the Translation Test, the Productive Levels Test and the Lex-30). Although they seemingly address the same type of knowledge, Fitzpatrick suggests that these tests are likely to differ, both from each other as well as from receptive tests, noting that the comparisons of the three tests generally produce only modest correlations.

2.4.2.2. Statistical Analysis of Free Production. An example of Laufer and Paribakht's (1998) free productive knowledge can be seen in the statistical analysis of free production in speech and writing. This type of measurement involves analysing the quality of the words that learners use productively in writing and spontaneous speech. Unlike the translation and elicitation methods, analysis of free production has learners produce language in a communicative manner. It thus offers the advantage of measuring vocabulary knowledge in a fitness for purpose context (Milton, 2009). Quality in these measurements is usually assessed based on lexical richness, which is a term used to describe the number of different words used in a given text and the diversity of the vocabulary (Torruella & Capsada, 2013). According to Daller and Xue (2007), lexical richness encompasses aspects such as *lexical diversity*, *lexical sophistication* and *lexical density*. Other researchers, such as Read (2000) have also included the *proportion of errors*. Each of these four aspects will now be discussed in detail.

2.4.2.2.1. Lexical Diversity. Lexical diversity was defined by Carroll (1938, p. 379) as "the relative amount of repetitiveness or the relative variety in vocabulary". In other words, it is the variety of active vocabulary that the learner uses in their speech or writing (Malvern & Richards, 2002). It has traditionally been assessed using the Type-Token Ratio (TTR) (Johnson, 1944), which shows the tokens (the number of words in a text) in relation to the types (the number of different words). For example, in the sentence discussed above, *The cat sat on the mat*, there are five types and six tokens, resulting in a TTR of 5:6 or 0.833 (Milton, 2009). The TTR has been used extensively to measure lexical diversity in child language research; however, it has often failed to discriminate between children at very different stage of language development (Richards, 1987). For example, one study carried out by Templin (1957, p. 115) investigated the language development of 480 children aged between three and eight and found that the ratio showed "little variation over the age range tested and among subsamples, sex, and SES groups". With regards to SLA, Perry (1998) analysed written essays of English language learners in terms of their TTR to determine if there was a relationship between the TTRs and the overall quality and marks received. However, results found no correlation between the lexical diversity ratio and high quality in written performance. As is pointed out by Milton (2009), the TTR has been criticised as being unreliable, in particular due to its sensitivity to length. This issue is explained by McCarthy and Jarvis (2007, p. 460), who state that "the more words (tokens) a text has, the less likely it is that new words

(types) will occur". This has prompted new measures which aim to measure lexical diversity while overcoming the issues regarding length, such as Guiraud's (1954) Index, Maas' (1972) Index, Malvern and Richards' (1997) D-measure, Daller et al.'s (2003) Advanced Guiraud Index, McCarthy's (2005) Measure of Textual Lexical Diversity and McCarthy and Jarvis' (2007) HD-D. In comparing the newer measures Measure of Textual Lexical Diversity and HD-D to the earlier measures D and Maas, Treffers-Daller (2013) reported that, although issues regarding segment size remain problematic, these new measures were found to be valid proxies for language ability. In addition, such measures have been shown to be very useful in the automatic analysis of students' vocabulary in writing tasks (Treffers-Daller, Parslow & Williams, 2018).

2.4.2.2.2. Lexical Sophistication. Lexical sophistication is described as the number of low-frequency words used by the learners (Daller & Xue, 2007), which are considered to be more advanced and sophisticated than high-frequency words (Vermeer, 2004). As is explained by Milton (2009, p. 135), "as learners improve in level and fluency, they are likely to increase the proportion of infrequent words they use in production", thus a higher number of low-frequency words demonstrates a higher language level. He highlights the difference between a learner producing the sentence "The cat sat on the mat" compared with the sentence "The feline reposed on the antique Persian rug" which, although has essentially the same meaning, is clearly very different with regards to productive capability.

Numerous programs have been developed to analyse lexical sophistication, such as Laufer and Nation's (1995) Lexical Frequency Profile (LFP), Meara and Bell's (2001) P_Lex, Heatley, Nation and Coxhead's (2002) RANGE, Meara and Miralpeix's (2004) V_Size, and Lindqvist, Bardel and Gudmundson's (2011) Lexical Oral Production Profile (LOPP). These programs work by comparing the given text against a basis of frequency bands based on written or spoken language corpora. Each one will now be discussed in detail, highlighting their key characteristics and findings from studies.

The LFP can be accessed in LexTutor via the programme VocabProfile (Cobb, 2018). Using Nation's (1984) word lists, it divides words in a text into four different frequency layers: the first 1000 most frequent word, the second 1000 most frequent words, the words in the AWL, and words not found in any of the previous three groups. It was introduced by Laufer and Nation (1995) in a study which analysed the lexical sophistication of 65 English learners of various L1s, who were grouped into three

different levels of proficiency. Each participant produced two written compositions of 300-350 words each in the same week. They also completed the VLT in order to measure their vocabulary size. Results found that the LFP obtained a reliable and stable measure of lexical richness across the two pieces of writing by the same participant, that it discriminated between participants of different proficiency levels, and that there was a correlation between the results of the LFP and vocabulary size. These findings were later questioned by Meara (2005), who used a series of simulations to test the claims made by Laufer and Nation and argued that the LFP was not all that sensitive in picking up modest changes in vocabulary size and that it worked best when comparing groups with very disparate vocabulary sizes. However, the LFP has been used to investigate lexical sophistication in writing with a variety of different purposes, for example, in analysing the first and final drafts of written compositions using the methodological approach of process writing (Muncie, 2002), determining whether using a bilingual dictionary improves quality of writing (East, 2006), and in a longitudinal study examining lexical use from a Complex Dynamic Systems perspective (Zheng, 2016). A French version of the LFP has also been validated by Ovtcharov, Cobb and Halter (2006), who analysed its capacity to distinguish between levels of lexical richness in oral data. They analysed oral interactions of 48 Anglophones learning French as an L2 and found significant differences in the frequency data of learners at different proficiency levels. However, as is noted by Lindqvist, Gudmundson and Bardel (2013), the appropriateness of using written corpora to analyse oral data is questionable. This led to the creation of Lindqvist et al.'s (2011) LOPP, discussed below.

Meara and Bell's (2001) P_Lex was presented as a mathematically more sophisticated, alternative approach to analysing short texts in terms of lexical complexity. Using Nation's (1984) word lists, it works by dividing the text into 10-word segments and analysing the number of frequent/easy and infrequent/hard words in each segment. Any words outside the most frequent 1000 words and proper nouns, numbers and geographical derivatives are categorised as infrequent. It then calculates the number of segments containing zero, one, two, and so forth infrequent words to obtain the P_Lex profile. In its initial presentation the authors reported the results of a study similar to that carried out by Laufer and Nation (1995). Forty-nine English learners of various L1s with various language proficiencies completed two pieces of written work of around 300 words in the same week and using the same titles as in Laufer and Nation's study. The participants also took the VLT to measure vocabulary size. The results suggested that

P_Lex worked well with shorter tests, making it particularly suitable in analysing compositions by low-level learners. This finding has been supported by Lu, Bao and Cui (2013) who, when comparing LFP with P_Lex found that although both measures demonstrated a certain degree of reliability, LFP was more likely to be affected by text length than P_Lex.

Heatley et al.'s (2002) RANGE, which is also accessible in LexTutor (Cobb, 2020), can produce different types of reports. One is a four-level report following the same criteria as the LFP, using the GSL/AWL lists. Another is a five-level report consisting of the first fourteen 1000 word divisions in BNC and a fifth level containing proper nouns and names. It can also be used with the BNC/COCA lists for 25,000 words. Kormos (2011) used RANGE in her investigation of linguistic and discourse characteristics in the written narratives of upper-intermediate FL learners in a Hungarian-English bilingual secondary school, and a group of native English speakers. Results showed major differences between the L1 and FL groups in relation to lexical sophistication. Daller and Hue (2009) have also used RANGE in their study of the vocabulary knowledge and academic success of 23 Chinese students attending a British university. Students completed a written essay as well as a C-test, and had also previously taken the IELTS exam. RANGE's output produced an LFP for each text, which was in turn used to compute a second measure of lexical sophistication, Guiraud Advanced (advanced types/square root of tokens). Results found that the LFP correlated significantly with Guiraud Advanced, that there were significant correlations between the C-test and the measures of lexical sophistication, and that IELTS and the C-test were useful predictor variables for the students' academic success and could anticipate poor performance in modules taken while abroad.

Meara and Miralpeix's (2004) V_Size analyses lexical sophistication using the assumption that texts created using a vocabulary of a particular size tend to have a characteristic shape: a text written by someone with a very small vocabulary will tend to contain only very frequent words while a text written by someone with a very large vocabulary will tend to contain fewer very frequent words and more infrequent ones. Using five different bands (Bands A, B, C and D containing the first, second, third and fourth 500 most frequent words in English, respectively, and Band E containing the remaining, infrequent words), a profile is generated for the selected text. As compared with other tools with a similar purpose, V_Size "goes beyond the mere shape of the profile generated by a text, and asks what the profile tells us about the size of the productive

vocabulary of the person who produced that text” (Meara & Miralpeix, 2016, p. 130). As is indicated by Moreno Espinosa and Agustín Llach (2010), V_Size has been used in the Spanish context with secondary school learners by Miralpeix (2008), who investigated whether there were differences in the number of words used by learners in different types of oral (interview, storytelling, roleplay) and written (composition) tasks. Results estimated that learners had vocabularies between 1,000 and 1,500 words, with a slight variation in the number of words depending on the task. While acknowledging that the idea behind V_Size is an interesting one, Roghani and Milton (2017) state that the scores produced are quite erratic and call for further research to demonstrate its reliability.

Finally, regarding spoken data, Lindqvist et al.’s (2011) LOPP was designed specifically to investigate advanced French and Italian L2 learners’ lexical profiles in oral production data. It has been used in a number of studies by these three authors to investigate lexical frequency profiles of learners at different proficiency levels (Lindqvist et al., 2011), thematic vocabulary and cognates among the low-frequency words used by learners at different proficiency levels (Bardel & Lindqvist, 2011), and a comparison of the original LOPP method with one that incorporates a distinction between advanced and basic vocabulary according to teacher judgements (Bardel, Gudmundson & Lindqvist, 2012). Lindqvist et al. (2013) provide a comprehensive overview of the tool and its usefulness in analysing oral production data.

2.4.2.2.3. Lexical Density. The concept of lexical density was first introduced by Ure (1971) and is a measure of the proportion of semantically full, or lexical words, as opposed to function words (Lindqvist et al., 2013). It may also involve calculating the noun density (the number of nouns divided by the total number of tokens) or analysing the number of verbs, adjective or adverb types per total lexical words (Johansson, 2009). The concept was further refined by Halliday (1985), who highlighted the importance of distinguishing between lexical items and grammatical items. For example, for Ure (1971) a phrasal verb such as *turn up* was counted as one lexical item (*turn*) and one grammatical item (*up*). Halliday, however, counts *turn up* as one lexical item, arguing that an item “function[s] in lexical sets not grammatical systems” (Halliday, 1985, p. 63). The early studies carried out by Ure (1971) and Halliday (1985) both outlined that lexical density was found to be lower for spoken language than for written. According to Ure’s (1971) study, lexical density in written English is typically higher than 40%, while in spoken English it is generally below 40%. This difference is exemplified in the following two

examples, the first of which Halliday (1985, p. 61) remarks is much more typical of spoken language:

1. If you invest in a rail facility, this implies that you are going to be committed for a long time. [lexical density = 0.35]
2. Investment in a rail facility implies a long-term commitment. [lexical density = 0.7]

On comparing the lexical density in the written work of native speakers and foreign students, Linnarud (1976) found that in most cases the non-native speakers' lexical density (38.66% to 47.15%) was below the predicted level for native speakers (48.54% to 50.45%). With regards to the relationship between lexical density and writing quality, Engber (1995) analysed 66 written essays by students with mixed language backgrounds in the intermediate to advanced range of an intensive English program. Quality scores were compared to four lexical richness measures, one of which was lexical density. Results found low, non-significant correlations between lexical density and quality.

2.4.2.2.4. Proportion of Errors. Finally, the proportion of errors, or number of errors, usually demonstrates how effective vocabulary use is in a language production (Goh, 2017). It has been defined as the “proportion and qualitative characteristics of lexical errors which include, among others, spelling mistakes, false friends, and other errors resulting from L1 interference” (Krzemińska-Adamek, 2016, p. 186). Although use of words which do not exist in the TL is evidently an important aspect of lexical richness, it is often not accounted for in lexical profiling analyses (Lindqvist et al., 2013). In Engber's (1995) study of intermediate learners, mentioned above, results found that there was a moderate significant correlation between the proportion of error and the test score. As can be expected, as score increased, the proportion of error decreased. Linnarud (1986), however, found a low, non-significant correlation between proportion of error and quality in the compositions of advanced learners. Engber (1995) attributes this to the difference in language level, suggesting that errors made by advanced learners may be less likely to impede comprehension and therefore may not have been as resulted in a lower quality score.

2.4.2.3. Association Tests. A word-association test is defined as “a psychological test in which the person being tested responds to a given word with the first word (or the first word in a specified category, such as an antonym) brought to mind” (“word-association test”, n.d.). It has been suggested that the use of association tests may help to overcome the shortfalls of the previous two measures of productive vocabulary knowledge: they can be standardised and provide a measure of overall productive ability, while at the same time allowing the learner a relative degree of freedom (Milton, 2009). While they are primarily used in the investigation of the mental lexicon, they are also beneficial in word-centred studies of depth of word knowledge (Dóczy & Kormos, 2016).

Word association tasks, which are the basis for investigating the lexical network, involve presenting the learner with a set of stimulus words one-by-one, and asking them to produce the first word that comes to mind in response (Read, 2004). For example, Meara and Fitzpatrick’s (2000) Lex30 provides the testee with 30 stimulus words and asks them to produce four association responses. A related task based on the principle of word association, with a focus on receptive knowledge, is the Word Associates Tests (Read, 1993, 1998), which has learners select from six to eight responses to a stimulus rather than supplying them. Notably, while some of the options are not related to the stimulus, some are related, but in different ways such as paradigmatic, syntagmatic and analytic. One clear advantage of such tasks is that the learner will not be hindered by the necessity to produce grammatical or well-structured language, but instead can focus solely on producing vocabulary (Milton, 2009). Furthermore, while initial responses are usually a frequent word, the subsequent responses are much more likely to be low frequency words, which can provide very informative results (Milton, 2009).

Since its first implementation two decades ago, Lex30 has been used with learners of a range of proficiency levels (Fitzpatrick & Meara, 2004; Walters, 2012), as well as in a wide range of educational contexts and ages, such as with university students (Fitzpatrick & Clenton 2010), secondary school CLIL students (Alejo & Piquer Píriz, 2016a) and young learners in FL contexts (Jiménez Catalán & Moreno Espinosa, 2005). Fitzpatrick and Meara (2004) investigated the reliability of the test using a test-retest study and two concurrent validity measures, one with native speakers and another with a set of collateral tests. Results showed the frequency profiles in the test-retest were broadly the same, that on average the native speakers scored higher than all but the most proficient non-native speakers, and that elicited vocabulary was measured with some accuracy. From these results they conclude that Lex30 operates with a degree of validity and

produces reliable results, however, they suggest that using a more up-to-date set of frequency bands may improve its accuracy. They also highlight that while Lex30 can certainly be used to determine whether a word is known in a “recall” sense, there is no indication that the participant can also actually “use” it. Walters (2012) compared L2 learners at different proficiency levels with the aim of determining the ability of the test to distinguish between different language levels. Concurrent validity was explored by comparing the participants’ results in Lex30 with two other productive vocabulary tests (the Productive Levels Test and a L1-L2 translation test). The issue raised by Fitzpatrick and Meara (2004) regarding whether Lex30 measures productive vocabulary or only recall was also addressed. The results support those outlined by Fitzpatrick and Meara (2004), with learners producing similar scores on parallel forms of the test and native speakers generally obtaining higher Lex30 scores than L2 learners, with the exception of some advanced learners. Results also found a significant correlation between Lex30 scores and other productive vocabulary tests. Walters concludes that the test is a reliable and valid productive vocabulary measure, however, suggested that “although the Lex30 test may be a valid test of productive vocabulary *use* for higher proficiency students, it is more valid as a test of productive vocabulary *recall* at the lower levels” (Walters, 2012, p. 183).

Regarding university students, Fitzpatrick and Clenton (2010) carried out further analysis of Lex30’s reliability and validity. In particular, they investigated the reliability of parallel forms, its internal consistency, its ability to reflect improvements in vocabulary knowledge after a six-week period of teaching intervention, correlations between the scores in similar tests and the differences in performance in speaking and writing. Results found that Lex30 produced consistent scores from the participants over the six-week period and that there was a significant improvement in the scores after teaching intervention, that a calculation of Cronbach’s alpha produced an acceptable internal consistency with a result of .866, that similar scores were obtained on parallel forms of the test, and that native speakers obtained higher scores than the learners with the exception of some advanced learners. They also found that Lex30 scores correlated significantly with the other productive vocabulary tests and that there was no significant difference between speaking and writing. Regarding this last finding, the authors also note that the correlation between the two forms was low, thus it may be questioned whether the vocabulary produced in one test would be the same as in the other. They suggest that the results present a strong argument for the test’s validity and advocate its

further investigation.

As for secondary school CLIL students, Alejo González and Piquer Píriz (2016a) assessed Lex30 with a particular focus on its reliability in testing children and teenagers (due to the potential lack of cognitive maturity affecting their attention span), correlation with general language proficiency, its measurement of vocabulary growth over a period of 18 months, and its sensitivity to the possible effect of the context of learning on the learners' productive vocabulary. The results found that although reliability scores were not very high, the students did not perform significantly less satisfactorily in the second half of the test; and that there were statistically significant moderate correlations between the participants' Lex30 scores and the three other L2 proficiency measures. Regarding vocabulary growth over time, results found that while there was statistically significant growth in the productive vocabulary of lower-level students, this was not the case for higher-level students; indicating a possible ceiling effect. Finally, results showed that the only vocabulary group in which the number of words produced significantly decreased between the two testing points was the low-frequency group. The authors attribute this to the learning context, highlighting that given the focus on specialised vocabulary in CLIL teaching, the frequency model could be disrupted as the learners' knowledge of technical vocabulary may cause inflated, less consistent results. The authors thus conclude that Lex30 may be an appropriate test for secondary school students, however, that scores should be interpreted with caution in specific educational contexts such as CLIL.

Finally, regarding young learners, Jiménez Catalán and Moreno Espinosa (2005) assessed the use of Lex30 in Spanish EFL learners in 4th grade of primary school, with the aim of determining whether it was a feasible instrument for measuring the L2 productive vocabulary of these young learners. Prior to the results, the authors highlight some methodological issues with Lex30, such as its inadequate procedure for dealing with repeated words, dialectal synonyms, acronyms which include dots, and misspelt words. They highlight that, as a result, Lex30 is an exploratory tool which is still in need of improvement. However, they note that the results found a significantly significant correlation with the receptive VLT and indicated that the index produced at Level 1 and Level 2 could be a feasible one, with decreasingly scalable results as the infrequency of occurrence of words increases.

As can be seen, these tests have by and large reported the validity of Lex30, albeit with suggestions for its improvement and some advising caution in the interpretation of the results within particular contexts. One particular example of word association test is

the lexical availability task. Given that it is of central focus in this project, LA research, which is based on association tasks (Jiménez Catalán & Dewaele, 2017), will be discussed in detail in Section 2.5.

2.4.2.4. Measures of Automaticity. Automaticity is clearly a vital aspect of developing both receptive and productive vocabulary skills (Henriksen, 1999), and researchers have consequently long argued for the need to focus on it (Meara 1996). Measures of automaticity of access, also referred to as fluency (Laufer & Goldstein, 2004), are generally concerned with how quickly a word can be recognised or accessed, and involve timed lexical decision tasks (Milton, 2009). With regard to receptive skills, fluency may mean the adequate recognition or comprehension speed when reading or listening, whereas with regard to productive skills, it may mean the adequate retrieval or production speed when speaking or writing (Schmitt, 2010). For example, subjects may be timed doing a task where they must assign an aspect of a meaning, such as living or non-living, to a stimulus word, or where they must identify a word in a string of random letters. As is pointed out by Milton (2009), although these tests seem to deal with receptive rather than productive skills, they appear to be testing a quality of knowledge which directly affects learners' productive vocabulary.

Segalowitz and Freed (2004) and Segalowitz et al. (2004) compared native-English learners of Spanish in a longitudinal study in two different contexts: study abroad and a formal classroom setting in their home university. These studies used an Oral Proficiency Interview to collect a pre- and post-test corpus of oral data and assessed lexical access using a semantic classification task whereby participants made speeded, two-alternative forced choice animacy judgements about single nouns (e.g., the boy = living; a boat = non-living). In Segalowitz and Freed (2004), participants were assessed on oral proficiency, oral fluency, grammar, vocabulary, pronunciation and communication strategies. Results showed that there was a correlation between lexical access speed and the proportion of filler-free speech, indicating that as lexical access improves, learners may become less reliant on fillers. Segalowitz et al. (2004) focused on oral performance gains in oral fluency as measured by temporal and hesitation phenomena, and their relationship with language contact and L2-specific cognitive measures such as lexical access, automaticity of lexical access and speed and efficiency of attention control. Results found significant interaction effects between oral proficiency, cognitive abilities and language contact.

Another study on the topic of lexical knowledge and spoken fluency was Hilton (2008), which investigated the basic measures of temporal L2 fluency in a corpus of oral productions in three different L2s: English, French and Italian. Data was analysed in terms of oral fluency measures such as speech rate, mean length of run, mean length of hesitation, percentage of production time spent hesitating, average hesitation times at various locations in the speech stream, rates of hesitation and retracing; and indicators of spoken performance such as mean length of utterance and error rate. Results found that there were positive correlations between language knowledge and spoken fluency, and negative correlations between language knowledge and negative production measures such as indicators of hesitation and error. The results also indicated that the lack of lexical knowledge or access to this knowledge appears to be behind the most serious disfluencies in the corpus, highlighting the vital role of lexical competence in spoken fluency.

The relationship between English vocabulary and verbal fluency has also been assessed by Portocarrero, Burrigh and Donovan (2007), who investigated the differences between monolingual and bilingual college students, and the relationship between the bilinguals' age of arrival in the U.S. with their English vocabulary. The Controlled Word Association Test was used to measure phonetic and semantic fluency, the former using the letters "F," "A," and "S." and the latter having participants name words which belonged to the categories *Animals*, *Kitchen* and *Actions*. Receptive and expressive vocabularies were also assessed using the Peabody Picture Vocabulary Test-III and the Expressive Vocabulary Test. Both groups had similar results in terms of phonetic fluency, however, the monolingual group outperformed the bilinguals in terms of semantic fluency. Bilinguals performance on English vocabulary was in the average range and they were found to have lower receptive and expressive English vocabularies than their monolingual peers. There was a moderate correlation between age of arrival and English vocabulary scores, with those who arrived at a younger age performing better.

Finally, Zhang and Lu (2014) carried out a longitudinal study investigating vocabulary breadth knowledge, by means of a paper-format version of the VLT, and vocabulary fluency development, by means of the same test in computer format to assess speed of meaning recognition. Results showed a significant effect of frequency level on both vocabulary breadth knowledge and vocabulary fluency at both testing points and for the rates at which the participants' vocabulary breadth knowledge and vocabulary fluency developed. It was also found that there was a weak relationship between vocabulary breadth knowledge and vocabulary fluency, the strength of which was affected by

frequency level. The authors also suggest that vocabulary fluency development falls behind vocabulary breadth knowledge growth, as is suggested by Laufer and Nation (2001).

As can be seen, studies investigating vocabulary and automaticity of access have used a wide range of measures including semantic classification tasks, analyses of various oral fluency measures, the Controlled Word Association Test, and a computer format of the VLT. This issue is pointed out by Milton (2009), who highlights the lack of established and fixed methodology in the measurement of automaticity, and the consequent array of measuring implements that have been used. However, despite the lack of consistency in testing measures, these studies largely indicate the importance of vocabulary knowledge in verbal and semantic fluency, given the necessity of quick lexical access for competent productive language.

2.5. Lexical Availability

This final section of the chapter will address an area of productive vocabulary that is of central importance to this project: lexical availability (LA). It will open with an overview of LA and be followed first by an explanation of the lexical availability task and then an overview of its use in previous research.

2.5.1. An Overview of Lexical Availability

As is pointed out by Jiménez Catalán (2014, p. v), “lexical availability is understood as the words that people have in their minds and that emerge in response to cue words that stand for domains closely related to daily life such as ‘Food and drink’, ‘Animals’, ‘Politics’, or ‘Poverty’”. Dating back to the 1950s, LA studies were first carried out in France, with the aim of teaching the French language to the people who made up the federation of territories known as *Union Française* (López Morales, 2014). When selecting the words which should be taught, priority was initially given to the most frequent words, as they were thought to be the most useful and the most used. However, it soon became clear that using frequency as the sole selection meant that many well-known words which were considered common by native speakers did not appear in the lists (Payne, 2016). For example, while words such as *t-shirt* or *teeth* are not very frequent, this does not necessarily mean that they are not important in everyday language. As Payne (2016) explains, LA studies emerged in order to compensate for this shortcoming in word frequency data, by foregrounding useful native-speaker vocabulary that had been excluded from frequency counts.

Michéa (1953) was the first to distinguish between “frequent words” and “available words”. As opposed to words which appeared frequently in the language, LA was understood as “the vocabulary flow *usable* in a given communicative situation” (López Morales, 2014, p. 3). As is explained by Fernández Smith, Sánchez-Saus Laserna and Escoriza Morerer (2012), this is based on an important distinction between thematic words and non-thematic words: while lexical frequency is made up of words which may appear in any text regardless of its subject (non-thematic words), LA is formed by the words that speakers have in their mental lexicon and whose use is determined by the specific theme of the text (thematic words). Thus, words which make up the “available lexicon” cannot be determined using frequency analyses as the available lexicon is only pertinent in actual lexical realisations and not potential ones (López Morales, 2014). This realisation was a turning point which led to the first LA studies carried out by Gougenheim, Michéa, Rivenc and Sauvageot (1964) in France, Dimitrijévic (1969) in Scotland, Mackey (1971) in Canada, and López Morales (1973) in Puerto Rico, which included important contributions to the theoretical and methodological systematisation of LA research (Fernández Smith et al., 2012). In particular, within a Spanish-speaking context, López Morales (1973) developed the Panhispanic Project of Available Lexicon, which aimed to analyse the available lexicon of non-specialised adult speakers in their final year of school before starting a degree (Samper Hernández, 2014). This research has provided guidelines for carrying out LA tests, allowing numerous researchers to carry out investigations which are methodologically homogeneous, and thus allowing for the comparison of results across different research (Fernández Smith et al., 2012), and from different points of view such as cognition, pedagogy, and language learning.

From a cognitive perspective, Hernández-Muñoz, Izura and Ellis (2006) have explored several factors influencing LA in Spanish L1 including concept familiarity, typicality, imageability, age of acquisition, word frequency, and word length. Notably, they have found concept familiarity, typicality and age of acquisition to be significant predictors of LA. More recently, Ferreira, Garrido Moscoso and Guerra Rivera (2019) investigated this same issue in 60 Chilean students with advanced L2 English with regard to two lexical prompts: *Body Parts* and *Food and Drink*. For *Body Parts*, age of acquisition and familiarity correlated highly with LA, with age of acquisition being found to be a significant predictor of LA. For *Food and Drink*, there was a moderate correlation between familiarity and frequency and LA, with frequency being found to be a significant predictor of LA.

LA has also been explored from a pedagogical perspective, in particular as part of the Fondecyt project (Salcedo, Ferreira, del Valle, Cerda & Friz, 2014) which analysed the lexicon of 233 Pedagogy in Mathematics students in two universities in Chile. By analysing five cue words specifically related to the discipline, the research could identify lexical deficiencies preventing optimal advances in learning development. These findings are of significant pedagogical importance, as they highlight that “lexical availability is not only responsible for quantitative measurements but also deals with how a particular concept is understood, identifying deficiencies or difficulties in the way words are interrelated” (Salcedo, del Valle, Quintanilla & Zambrano, 2016, p. 3676).

Finally, the importance of this research within the context of language learning cannot be overlooked. Fernández Smith et al. (2012), for example, highlight the numerous benefits of LA tests for language teaching and learning:

they allow us to analyse the vocabulary known by foreign students, to detect the possible differences between the speakers of different mother tongues, to observe the learning phases, to compare their available lexicon with that of other communities of native speakers, and to provide the editor of teaching manuals and books with teaching material that is suitable for each of the different levels of teaching. (p. 15)

2.5.2. The Lexical Availability Task

Most LA studies to date have benefitted from methodological homogeneity, using the Lexical Availability Task (LAT) as a data collection instrument. As is outlined by Samper Hernández and Jiménez Catalán (2014), the task is generally carried out by means of a paper-and-pencil questionnaire. Participants are presented with cue words (also known as “prompts”, “centres of interest”, or “semantic categories”) and asked to write down all words that come to their mind in response. Each category is displayed on a different page with numbered lines, and participants are given two minutes for each cue word and cannot move to the next category until after this two-minute period.

With regard to the cue words, most studies have used the 16 categories proposed by Gougenheim et al. (1964) in the first LATs: *Parts of the human body, Clothing, Parts of the house, House furniture, Food and drink, Objects on the table for the meal, The kitchen and its utensils, School furniture and materials, Heating and lighting, The city, The countryside, Means of transport, Farm and garden work, Animals, Games and entertainment, Jobs and professions* (Šifrar Kalan, 2015, p. 197). While the number of

cue words which are chosen may vary, using similar categories and methodological approach has allowed for cross-comparison of results in a wide range of studies.

The systematicity in the results has led researchers to conclude that it is a reliable task which produces consistent results, both in testing the L1 as well as EFL available lexicon (Canga Alonso, 2017). Furthermore, as is highlighted by Jiménez Catalán and Fitzpatrick (2014), vocabulary that is retrieved by means of the LAT provides a useful indication of the lexical resources which are available to learners, and importantly, it obtains this rich set of data in a rather economical way.

2.5.3. Previous Research on Lexical Availability

As was noted for vocabulary in general, Jiménez Catalán (2014, p. v) has highlighted that, despite an awareness of the importance of LA for lexical and communicative competence, “little research has been conducted on this issue in second or foreign language education, and practically nothing has been done in the field of vocabulary studies”. Research is particularly sparse in L2 English, with a clear focus to date on L2 Spanish (Martínez-Adrián & Gallardo-del-Puerto, 2017). However, recent years have seen an increase in interest and research in this area, leading to a number of important studies and findings, particularly with regards to EFL or FFL within a Spanish context, as detailed in Table 2.6 and explained below (for an overview of LA studies dealing with Spanish, see Samper Hernández & Jiménez Catalán, 2014).

Firstly, with regard to young learners, two studies were carried out by Jiménez Catalán and Ojeda Alba, investigating the LA of males and females (2009a), and of CLIL and non-CLIL students (2009b), in learners in 6th grade. Participants in the first study included 210 learners, 105 male and 105 female, while those in the second included 86 learners, 42 CLIL and 44 non-CLIL. Both studies included 15 different prompts. Regarding gender, results of the first study found statistically significant differences in the average number of words produced by the two groups, with girls achieving higher means than boys in all 15 prompts of the LAT. Regarding CLIL instruction, the researchers found that the non-CLIL group performed significantly better on a language placement test than the CLIL group, and that although they also produced a higher number of words in the LAT than the CLIL group, the difference was not statistically significant.

Table 2.6*Key Foreign Language LA Studies in Spain*

Authors	Age group	Focus	Main Findings
Jiménez Catalán & Ojeda Alba (2009a)	Young learners (6 th grade)	Gender	Statistically significant differences between genders: girls achieved higher means than boys in all fifteen prompts of the LAT.
Jiménez Catalán & Ojeda Alba (2009b)		CLIL vs. non-CLIL	The non-CLIL group did significantly better on a language placement test than the CLIL group and produced a higher number of words in the LAT than the CLIL group (no statistically significant difference).
Fernández Fontecha (2010)	Teenagers (8 th grade)	Gender and Motivation	Statistically significant differences between the two groups in favour of the female learners.
Jiménez Catalán & Fitzpatrick (2014)	Young learners (6 th grade) Teenagers (8 th grade)	Age and Word frequency	Older students produced more tokens and types than younger learners; lexical retrieval depended on the cue words; qualitative differences were found between each group.
Agustín Llach & Fernández Fontecha (2014)	Young learners (6 th grade) Teenagers (9 th grade)	Gender and Age	Statistically significant difference between genders at both points: girls produced a higher number of words; similarities across genders in most and least productive prompts.
Jiménez Catalán et al. (2014)	Young learners (6 th grade) Adults (1 st year university)	Age	No statistically significant difference between the two groups, though the adult group produced a higher number of words; shared and non-shared vocabulary in both groups.

Table 2.6 (continued)

Gallardo del Puerto & Martínez Adrián (2014)	Adults	True vs. False beginners	False beginners produced a higher number of words than the true beginners; adults' performance was similar to young learners in other studies.
Fernández Orío & Jiménez Catalán (2015)	Teenagers (10 th grade)	EOI vs. EFL	Statistically significant differences between the two groups in favour of those attending the Official Language School.
Santos Díaz (2015, 2017a, 2017b, 2017c)	Adults	L1 Spanish L2 English or French	Participants produced the highest number of words in their L1; higher mean in the English LAT than the French LAT.
de la Maya Retamar (2016)	Teenagers (8 th + 9 th grade)	Receptive and productive vocabulary and LA in French, gender and motivation	Motivation correlated with productive vocabulary but not with receptive vocabulary or LA; no gender differences.
Jiménez Catalán & Dewaele (2017)	Young learners (6 th grade)	Emotion words v non-emotion words	Participants retrieved a higher number of words for non-emotion prompts than for emotion prompts and for positive emotion prompts than negative emotion prompts.
Jiménez Catalán & Agustín Llach (2017)	Teenagers (8 th + 10 th grade)	CLIL vs. non-CLIL	The younger CLIL group outperformed both their non-CLIL peers and also the older non-CLIL group.

Table 2.6 (continued)

Canga Alonso (2017)	Teenagers (12 th grade)	Native speakers vs. EFL learners	Spanish L1 speakers produced a higher number of words; the prompts <i>Food and Drink</i> and <i>Countryside</i> were the most and least productive.
Jiménez Catalán & Fernández Fontecha (2019)	Teenagers (12 th grade)	L2 vs. L3	The L2 group produced both a greater number of words as well as a higher percentage of infrequent words in the most productive prompt.
Jiménez Catalán & Canga Alonso (2019)	Teenagers (12 th grade)	Gender	No statistically significant difference between genders; some qualitative differences in the types of words produced.

Jiménez Catalán and Dewaele (2017), which also analysed the LA of 6th grade students, focused specifically on the words produced in response to emotion prompts (*Love, Hate, Happy* and *Sad*) and non-emotion prompts (*School* and *Animals*). Results showed that participants produced a higher number of words for non-emotion prompts than for emotion prompts and for positive emotion prompts than negative emotion prompts. Predominance was also observed in the retrieval of nouns over other word classes. A number of studies have also compared young learners with teenagers or adults. For example, Jiménez Catalán and Fitzpatrick (2014) investigated the difference between learners in 6th and 8th grade in response to nine different cue words. They adopted an important new approach to analysing LA, by applying a word frequency framework, analysing the number of words produced and the proportion of infrequent to frequent words in each domain. Results found that the older group produced many more tokens and types than the younger learners, that the words retrieved by the two groups differed qualitatively, that vocabulary retrieval depended on the cue words used, that most lexical growth was seen at the 1K band, followed by the off-list band, and that little evidence was found of academic vocabulary. Another important study which compares young learners (6th grade) with teenagers (9th grade) is that by Agustín Llach and Fernández Fontecha (2014), which investigated the issue of gender using a longitudinal approach.

As in Jiménez Catalán and Ojeda Alba (2009a), results found that the female group produced significantly more responses than the male group, both in 6th grade and three years later in 9th grade. However, results showed that significance values decreased as learners grew older. They also showed that learners in 9th grade produced significantly more responses than in 6th grade, demonstrating that the learners' LA increased as they got older and received more exposure to the language. A final study dealing with young learners was carried out by Jiménez Catalán, Agustín Llach, Fernández Fontecha and Canga Alonso (2014) and compared the differences between 6th grade primary students and 1st year university students of the same language level. They investigated the number of words and type of words retrieved by the participants in response to the cue words *Town* and *Countryside*. Although the number of words produced by the adult group was higher, there was no statistically significant difference between the two groups. This was perhaps, as the authors suggest, due to the small size of the sample. Qualitatively, the results also showed the existence of both shared and non-shared vocabulary in both groups, as well as a predominance of nouns in the words retrieved.

Secondly, with regard to teenagers, Fernández Fontecha (2010) analysed gender differences in a cohort of 250 8th grade Spanish EFL learners (139 male, 111 female) in terms of LA and motivation (see [Section 4.3.2](#) for a discussion of the latter). Results of the LAT indicated that there was a statistically significant difference between the two groups in favour of the female learners, who retrieved a higher number of words. Other noteworthy research on the LA of teenage students has been the doctoral study carried out by de la Maya Retamar (2016). This work investigated for the first time Spanish students' LA in French in students in 8th and 9th grade ($n = 81$) by means of a receptive vocabulary test, productive vocabulary test and a LAT, as well as a motivation test. The tests, which were all adapted in order to test the students' command of French, included Meara and Milton's (2003) X-Lex test for receptive vocabulary, Meara and Fitzpatrick's (2000) Lex30 for productive vocabulary test and a 15-prompt LAT. A 23-item French language motivation questionnaire, used by Alejo González and Piquer Píriz (2016b) and based on Dörnyei, Csizér and Németh (2006), was also administered. Results found that students' grade had the greatest influence on the results, that motivation results correlated with productive vocabulary but not with receptive vocabulary or LA, and that there were no differences between the male and female groups in any of the vocabulary measures. Jiménez Catalán and Agustín Llach (2017), in order to deal with the issues regarding CLIL and time of exposure mentioned above, compared teenage CLIL and non-CLIL

learners after an equal number of hours of exposure to English by means of a LAT. Participants included 24 CLIL students and 26 non-CLIL students in 8th grade, as well 19 non-CLIL students in 10th grade with an equal number of hours of exposure as the 8th grade CLIL group. Results showed that the 8th grade CLIL group not only outperformed their non-CLIL peers, but also the 10th grade non-CLIL group. With regard to LATs with Spanish students in their last year of compulsory secondary education (10th grade), there has been just one published study carried out by Fernández Orío and Jiménez Catalán (2015). This research sought to determine whether there were differences between students in two different learning contexts: those in regular EFL classes and those participating in the English collaborative programme between the school and the Official Language School of the capital city of the region. Results showed significant quantitative differences between the two groups, in favour of those attending the Official Language School. Other studies dealing with older teenagers are Canga Alonso (2017), Jiménez Catalán and Fernández Fontecha (2019) and Jiménez Catalán and Canga Alonso (2019), which all investigated the English LA of 12th grade Spanish students. Canga Alonso (2017) aimed to compare the LA of native Spanish speakers and EFL students. Results showed that the Spanish L1 speakers produced a higher number of words, with the prompts *Food and Drink* and *Countryside* being the most and least productive, respectively. Jiménez Catalán and Fernández Fontecha (2019) aimed to determine whether there were differences between students learning English as a L2 (monolingual students learning English) as opposed to as a third language (L3) (bilingual students learning English). Results found that the L2 group produced both a greater number of words as well as a higher percentage of infrequent words in the most productive prompt. Jiménez Catalán and Canga Alonso (2019) compared the LA of 265 12th grade Spanish EFL learners in terms of gender (171 females and 94 males). In contrast to the other LA studies investigating LA in children and younger teenagers (Jiménez Catalán & Ojeda Alba, 2009a; Fernández Fontecha, 2010; Agustín Llach & Fernández Fontecha, 2014), no statistically significant difference was found between the two groups in the number of words produced. Qualitative analysis did, however, reveal differences among male and female students in the specific words used for some prompts. For example, in the prompt *Professions*, words such as “engineer”, “politician” and “professor” were used only by male participants while “singer”, “musician” and “shop assistant” were used only by female participants.

Finally, regarding adults, Gallardo del Puerto and Martínez Adrián (2014) offer

an analysis of senior learners' LA, comparing those with previous English contact with true beginners. Results showed that the false beginners produced a higher total number of words than the true beginners, and a higher number of words in most of the 15 semantic categories. They also reported striking similarity between the LA of the beginner adults and that of young learners in other studies, suggesting that beginners experience similar stages in vocabulary acquisition, regardless of age. While it takes a very different approach from the aforementioned studies, Santos Díaz's (2015, 2017a, 2017b, 2017c) work on LA is of particular interest given that it has investigated LA in both English and French as an L2. The data for the studies were collected from 171 students of a master's degree in Teaching Training. All participants completed a LAT which contained nine prompts, both in Spanish and in an FL (150 participants in English and 21 participants in French), and a sociological questionnaire. The results have been discussed in relation to a wide range of issues: the influence of knowing multiple FLs on LA (Santos Díaz, 2015), the influence of reading frequency on LA (Santos Díaz, 2017a), the semantic relations of words in Spanish L1, English L2, and French L2 (Santos Díaz, 2017b), and establishing guidelines for the selection of the available lexicon and to empirically test its suitability (2017c). This last study is of particular interest given that it provides the number of words retrieved in each language. In the nine categories, the participants produced a mean of 397.67 words in Spanish, 261.67 in English and 221.33 in French. The study also indicates the percentage of non-shared words in each language: 42.44% in Spanish, 44.76% in English and 50.09% in French. However, as is pointed out by Ferreira et al. (2019), the research does not relate LA to other factors in the way other research in the field has done. Although it was not carried out in a Spanish context, also noteworthy is a study by Šifrar Kalan (2014), which investigated the LA of 40 Slovenian university students (20 English learners and 20 Spanish learners) using eight semantic categories. Results showed similarities in the LA of the two groups, with both groups producing the highest number of tokens in the categories *Animals* and *Food & Drink*, while the categories *Games & Entertainment* and *School: Furniture & Material* were the least productive. The average number of words produced was slightly higher for the English group than for the Spanish group. In both groups, the highest means were found in the most advanced learners, suggesting a relationship between language proficiency and LA. Qualitatively, a high degree of similarity was observed across the semantic categories for the top ten available words, with a large number of shared words across most categories, and in the most

common semantic prototypes, indicating that the groups have similar lexical connections regardless of the FL at hand.

While the above studies have taken the first steps in investigating LA in Spanish learners of English and French, it is clear that much work is needed to further investigate this area of research. In particular, Canga Alonso (2017) has made a number of suggestions for future LA research. Firstly, he highlights the need to include some measure of proficiency alongside the LAT, so as to determine the effect of language level on the number of words retrieved. Secondly, with regard to students who are studying content classes via an FL, he suggests focusing on prompts which may be relevant to that subject, for example, investigating the prompt *Parts of the Body* with students who have learnt Natural Sciences and Biology through English.

Chapter 3: Motivation in Second Language Learning

This chapter opens with an introduction to motivation in L2 learning. It then provides a historical overview of the main theories and approaches which have shaped L2 motivation research. The third section offers an introduction to L2 motivation research, with the following three sections addressing motivation research in relation to three key issues in this thesis: the language of study, learning context, and gender.

3.1. Motivation in Second Language Acquisition

This opening section of the chapter will provide an introduction to the concept of motivation in SLA. It will first explain what is understood by motivation in language learning and why the concept has come to be so important, and then provide an overview of some key terminology used in L2 motivation research.

3.1.1. What is L2 Motivation and Why is it Important?

Motivation is a concept which, although somewhat intuitively comprehensible, has proven to be exceedingly complex in its definition and conceptualisation (Dörnyei & Ushioda, 2013). This is extremely apparent in Kleinginna and Kleinginna's (1981) categorisation of motivation definitions in the field of psychology alone, which compiled over 100 statements defining or criticising the concept in an effort to resolve the terminological confusion. One reason behind this disparity may be that, given that motivation research intends to answer the question "why do organisms think and behave as they do?" (Weiner, 2013, p. 1), it is unlikely that the complexity of such an issue can be explained by a single theory (Dörnyei & Ushioda, 2013). Furthermore, the concept of motivation will also be viewed differently depending on the theoretical background of the researcher. Murayama (2018), for example, highlights the lack of a unified and cross-disciplinary approach within and beyond the field of psychology, as well as the fundamentally different understanding and approach to the concept across fields- two factors which will evidently lead to further terminological confusion. According to Dörnyei and Ushioda (2013, p. 4), however, most researchers agree that motivation concerns the direction and magnitude of human behaviour, in other words: "*why* people decide to do something, *how long* they are willing to sustain the activity, and *how hard* they are going to pursue it".

Within the field of SLA, Gardner (1985) first defined motivation as the combination of four vital factors: effort to learn the language, having a goal, a desire to

achieve that goal, and a favourable attitude towards learning the language. Thus, a motivated language learner is seen as someone who works hard, wants to accomplish something, and has a positive disposition towards the language. As Cook and Singleton (2014) note, motivation means more than just feeling good about doing a certain task; it is whatever actually pushes the person to carry out the job. Thus, it is not simply the fact that someone likes a language, but rather that this affection will drive them to improve. In addition, effort alone does not indicate motivation: a motivated person spends effort towards the aim, but the person expending effort is not inevitably motivated (Gardner, 1985).

After over six decades of L2 motivation research, there is now a very clear consensus that motivation is a vital part of students' learning (Keblawi, 2009). As Dörnyei (2014) has indicated, it is so important that even the most exceptional language learners will fail to reach long-term goals if they lack the motivation to do so. Hadfield and Dörnyei (2013) also highlight the importance of motivation, pointing out that it is one of the most commonly used terms by students and teachers to explain success or failure in language learning. The reason why language learning motivation is so important has been attributed to a number of factors. According to Oxford and Shearin (1994), research has shown a direct influence of motivation on how often students use L2 learning strategies, how much they interact with native speakers, the amount of TL input they receive, their curriculum-related achievement tests, their general proficiency and how long they retain their L2 skills after language study has finished. In addition, as Cook and Singleton (2014) have noted, attitude and motivation are seen as crucial in any learning context, including L2 learning contexts, because they are seen as determining the extent of a learner's active involvement in learning. Active involvement, in turn, is suggested to be necessary for language acquisition to take place, with language teaching systems which provide greater active involvement reporting great success (Krashen, 1976).

3.1.2. Key Terms in L2 Motivation

Before outlining the key theories and approaches to L2 Motivation in the next section, an overview will be provided of some of the key terms and types of motivation which will be discussed in the sections to follow: orientation, integrativeness, instrumentality, intrinsic, extrinsic, amotivation and demotivation.

Firstly, one important term in L2 motivation research is the learner's *orientation*, defined by Gardner (2010, p. 16) as "an overall aim, purpose, direction, and/or goal of

the activity”. In other words, it is the reason why the L2 is being learned. According to Ortega (2014), orientations are not antithetical or mutually exclusive, and thus a learner may have several orientations at any given time. She highlights five types of orientations in language learning: instrumental, for knowledge, to facilitate travel, to foster friendship, and for integrative reasons (Ortega, 2014, p. 175). Thus, someone may learn a language so as to be able to read a book in that language (knowledge) while at the same time wanting to visit a country where that language is spoken (travel) or meet people who speak that language (friendship).

Out of these motivational substrates, known as antecedents (Ortega, 2014), research has generally focused on two theoretical concepts: *integrativeness* and *instrumentality*, with the former gathering most attention to date (Dörnyei & Ushioda, 2009). Integrativeness “reflects the individual’s willingness and interest in social interaction with members of other groups” (Gardner & MacIntyre, 1993, p. 159.). In other words, it concerns the extent to which the learner wants to be involved in the TL community. As indicated by Ortega (2014), it involves three dimensions:

1. A favourable attitude toward the TL speakers.
2. A general desire to learn FLs and low ethnocentrism.
3. An integrative orientation, that is, an endorsement of reasons for learning the language based on interaction with L2 members.

Thus, a high level of integrativeness would be characterised by an individual who admires the people who speak the L2, likes learning FLs and wants to study a language because it will enable them to meet and speak to people in that language. Instrumentality, on the other hand, is concerned with a desire to gain something practical from learning the language (Cook & Singleton, 2014). For example, a learner may want to learn a language simply to be able to pass an exam in school or to be able to get a particular job. While both integrative and instrumental motivational orientations have been found to be effective, an integrative orientation has been shown to be particularly effective in the long term (Hummel, 2014). However, as indicated by Cook and Singleton (2014), three qualifications should be made when discussing the distinction between integrative and instrumental motivation. Firstly, as noted above, the orientations are by no means mutually exclusive, and very often both are present. For example, the same learner may be motivated to learn a language both because of a desire to communicate with its speakers and also in order to facilitate future economic advancement. Secondly, the global

spread of English has led to considerable debate regarding the basic premise of the integrative concept (Dörnyei & Ushioda, 2009), which states that the learner “must be willing to identify with members of another ethnolinguistic group and take on very subtle aspects of their behaviour” (Gardner & Lambert, 1972, p. 135). Given the fact that non-native speakers of English now vastly outnumber native speakers (Jenkins, 2002), English learners may evidently focus on communication with speakers of different linguistic backgrounds (Breiteneder, 2005). It thus becomes increasingly difficult for learners to identify a clear target group or culture, making concepts such as integrativeness and attitudes toward the TL community increasingly unclear (Yashima, 2009). Some researchers such as Yashima (2009, p. 3) have consequently moved towards evaluating learners instead on their international posture, which is the “tendency to see oneself as connected to the international community”, rather than a specific L2 group. The global spread of English has also come with repercussions on learners’ instrumental motivation (Dörnyei & Ushioda, 2013), as given that English is increasingly seen as a basic educational skill, crucial for economic and professional advancement, there are likely to be fundamental differences in a learner’s motivation for learning English as opposed to other languages. Block and Cameron (2002), for example, have highlighted how language learning and communication skills which are necessary due to globalisation influence motivation towards instrumentality. A final qualification regarding these orientations is that not all reasons given by learners for wanting to master a particular L2 fit easily into Gardner’s framework. Cook and Singleton (2004, p. 97) highlight research carried out by Oxford and Shearin (1994) in which participants report reasons for studying Japanese such as intellectual curiosity, setting themselves a personal challenge, showing off to peers, having a fascination with the Japanese writing systems or acquiring a type of secret code.

Motivation can also be *intrinsic* or *extrinsic*. The former occurs “when the learning activity and the learning environment elicit motivation in the learner” while the latter “is defined in terms of motivation deriving from outside regulation” (Cook & Singleton, 2014, p. 100). For example, an intrinsically motivated learner may wish to become a competent speaker of the TL because of a personal desire to excel, whereas an extrinsically motivated individual may be more concerned with improving their language skills to get a higher paying or more prestigious job (Hummel, 2014). As indicated by Brown (2007, p. 175, referring to Bailey, 1986) and summarised in Table 3.1, both intrinsic and extrinsic motivation may be either integrative or instrumental.

Table 3.1*Motivational Dichotomies*

	Intrinsic	Extrinsic
Integrative	L2 learner wishes to integrate with L2 cultural group	Someone else wishes the learner to know L2 for integrative reasons
Instrumental	L2 learner wishes to achieve goals utilizing L2	External power wants L2 learner to learn L2 to achieve goals

Note. Adapted from *Principles of language learning and teaching* (5th ed. p. 175), by H. D. Brown, 2007, Pearson Longman. Copyright 2007 by Pearson Education.

As can be seen, an individual who is integratively and intrinsically motivated wishes to learn the TL in order to use it with those who speak it, for example, wishing to learn English so as to be able to make English-speaking friends. On the other hand, in the case of an integratively and extrinsically motivated individual, this desire is encouraged from an external individual, for example, a parent wishes for their child to learn the TL so that they may become integrated in the TL country. Instrumentally and intrinsically motivated learners are those who have a personal desire to learn the TL in order to achieve a particular goal, for example, learning the TL in order to advance in one's career. Finally, for instrumentally and extrinsically motivated individuals this drive again comes from an external source, for example, a parent wishes for their child to learn the TL so as to better prepare them for their economic future.

In addition to these terms which deal with the varying sources of motivation, a number of terms have also been used to denote a lack of motivation. *Amotivation*, for example, refers to the lack of any kind of motivation, either intrinsic or extrinsic (Dörnyei & Ushioda, 2013). *Demotivation*, on the other hand, generally entails some kind of conflict, resentment and disaffection (Cook & Singleton, 2014) or, as Zhang (2007, pp. 213-214) explains, is “the force that decreases students' energy to learn and/or the absence of the force that stimulates students to learn”. While both terms entail a lack of motivation, amotivation concerns unrealistic outcome expectations while demotivation concerns specific external causes (Dörnyei & Ushioda, 2013). For example, an amotivated language learner is someone who sees no point in learning the language, perhaps because they see it as an unobtainable goal or because they see no useful reason for knowing the language. On the other hand, a demotivated language learner is

understood to have once been motivated, but due to some external factor, such as an unpleasant teacher or bad grade, has become demotivated.

Bearing in mind the above terminology and distinctions, the following section will offer an overview of the key theories and approaches in the history of L2 motivation.

3.2. An Overview of L2 Learning Motivation: Key Theories and Approaches

The starting point of L2 motivation research is characterised by the work of Robert Gardner, whose thesis, under the supervision of Wallace Lambert, gave birth to the field of L2 motivation and led to decades of research on the topic (Al-Hoorie, 2017). Since research first began investigating L2 motivation over six decades ago, there have been a number of key periods, which Ushioda and Dörnyei (2012) identify as the following:

1. The social-psychological period (1959–1990)
2. The cognitive-situated period (the 1990s)
3. The process-oriented period (turn of the century)
4. The socio-dynamic period (current)

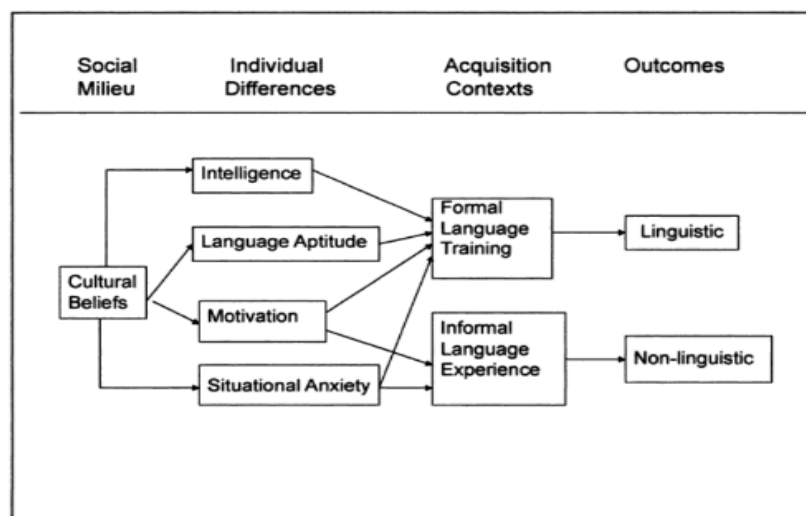
Each of these periods has introduced new theories and models for addressing motivation in SLA, with each one adapting their approach based on findings from the previous phase, as knowledge on the topic grew. It should, however, be noted that while these four periods are useful in understanding the main trends and approaches taken throughout the history of L2 motivation, the dates used are merely general indicators, as there has been a considerable amount of overlap between stages (Al-Hoorie, 2017). For example, some research from the cognitive-situated period retained elements from the social-psychological period, while some cognitive theories were still pursued in subsequent periods and even to this day. As a result, in the numerous historical overviews of L2 motivation (e.g., Dörnyei, 2005; Dörnyei & Ushioda, 2013; Ushioda & Dörnyei, 2012; Boo, Dörnyei & Ryan, 2015; Dörnyei & Ryan, 2015; Guerrero, 2015; Al-Hoorie, 2017; Woodrow, 2017), some earlier researchers have discussed three periods instead of four (leaving out the latest socio-dynamic period), have later on merged two periods into one stage (e.g., merging the process-oriented and socio-dynamic periods), or have discussed the same framework under the banner of a different period (e.g., the social constructivist model, which has been reviewed as part of both the cognitive-situated and process-oriented periods). With this caveat in mind, the following four sections will provide an overview of the four-period distinction outlined by Ushioda and Dörnyei (2012), outlining their main models, theories and contributions to L2 motivation.

3.2.1. The Social-Psychological Period

The starting point of L2 motivation research, the social-psychological period, is characterised by the work of Robert Gardner and his associates in Canada (Guerrero, 2015). The research in this period, which focused on macroperspective language learning, was based primarily on the assumption that there is a key difference between learning an L2 and other school subjects, given that the former has the additional requirement of openness towards the L2 group and a willingness to adopt different features of the language (Al-Hoorie, 2017). Of particular importance in this period is Gardner’s (1979) socio-educational model of motivation and L2 learning, which is likely the most frequently cited motivational framework in SLA research (Cook & Singleton, 2014). Although it first emerged in the 1970s, the model has been updated numerous times over the years. As Gardner (2019) highlights, Gardner and Smythe (1974) originally focused on a classification of scales with various motivational properties, including Group Specific Attitudes, Course Related Characteristics, Motivational Indices and Generalized Attitudes, while Gardner (1979) produced an updated model linking four interrelated variables. As shown in Figure 3.1, these variables included the social milieu (cultural beliefs), individual differences (intelligence, aptitude, motivation and situational anxiety), the context (formal or informal) and outcomes (linguistic or non-linguistic) (Gardner, 1985).

Figure 3.1

Gardner’s (1979) Socio-Educational Model



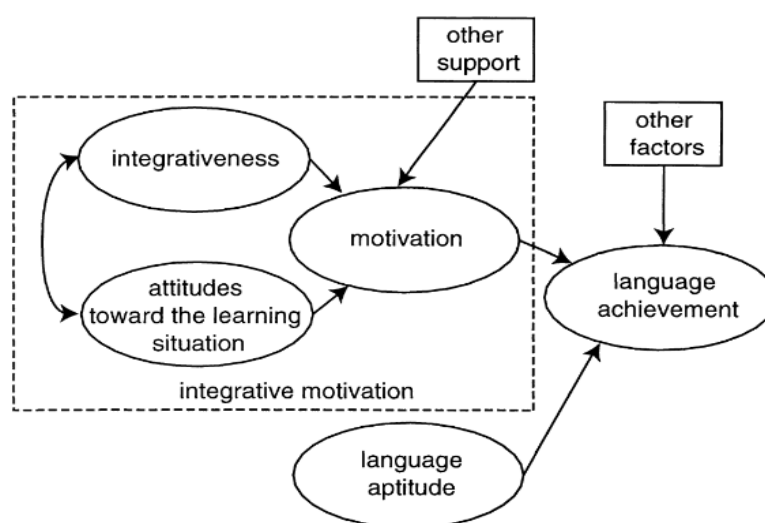
Note. From *Motivation and second language acquisition: The socio-educational model* (p. 83), by R. C. Gardner, 2010, Peter Lang. Copyright 2010 by Peter Lang Publishing.

In this model, attitudes were seen to influence motivation, which in turn influenced language achievement (Gardner, 2001). In addition, aptitude and motivation were seen to take primary position over other variables, with students with higher levels of language aptitude and motivation outperforming those with lower aptitude and motivation.

As shown in Figure 3.2, further revisions were introduced in Gardner (2000).

Figure 3.2

Gardner's (2000) Socio-Educational Model



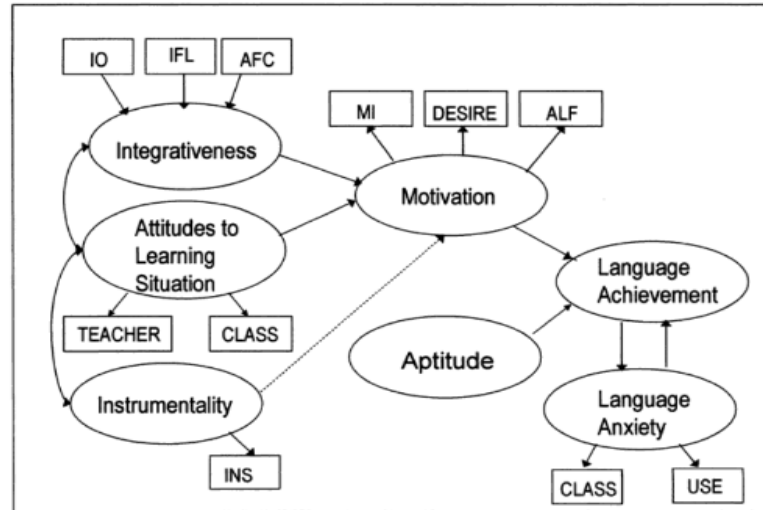
Note. From “Integrative Motivation and Second Language Acquisition” by R. C. Gardner, in Z. Dörnyei & R. Schmidt (Eds.), *Motivation and second language acquisition* (p. 5), 2001, University of Hawaii Press. Copyright 2001 by Second Language Teaching & Curriculum Center.

In this updated model, integrativeness and attitudes towards the learning situation were understood as two correlated variables which influenced the learner’s L2 motivation, while motivation and language aptitude in turn influenced language achievement (Gardner, 2001). According to Gardner (2001), this model maintains that if someone shows high levels of Integrativeness and/or very positive Attitudes Toward the Learning Situation, but these are not linked with Motivation to learn the language, these variables will not be particularly highly related to achievement. Similarly, a person who demonstrates high levels of Motivation which are not supported by high levels of Integrativeness and/or favourable Attitudes Toward the Learning Situation may not show these high levels of motivation consistently. Integrative Motivation represents a complex of these three variables.

Gardner (2006) later introduced a structural equation representation which linked the scales (Figure 3.3).

Figure 3.3

Gardner's (2006) Structural Equation Representation of the Socio-Educational Model



Note. From *Motivation and second language acquisition: The socio-educational model* (p. 88), by R. C. Gardner, 2010, Peter Lang. Copyright 2010 by Peter Lang Publishing.

As can be seen, motivation itself was measured via three different scales, which assessed its different facets: Motivational Intensity (MI), that is, the effort and persistence of the learner; Desire to Learn the Language (DESIRE), that is, the personal interest in the goal of learning the language; and Attitudes toward Learning the Foreign Language (ALF), that is, the affect experienced (Gardner, 2019). In addition, the 2006 model also described the learners' orientation in terms of the two classical categories Integrativeness (subdivided into Integrative Orientation [IO], Interest in Foreign Languages [IFL], and Attitudes toward the Language Community [AFC]) and Instrumentality (INS) (Doiz et al., 2014a). Though these orientations are solely motivational antecedents, rather than strictly a part of motivation, they have somewhat ironically become the most widely known concepts associated with Gardner's work in the field of L2 motivation (Dörnyei & Ushioda, 2013).

Finally, and most recently, Gardner (2010) presented an updated model based on research conducted in different countries on the acquisition of English as a global language. This adaptation, which involved producing an internationalised version of the model's testing instrument (discussed below), demonstrated the applicability of the model

to contexts other than the original Canadian context for which the first model was devised. Use of the model across 12 large samples in six different countries served to demonstrate substantial relationships between integrative motivation and indices of English language achievement (Gardner, 2010).

Research carried out using Gardner’s framework has been based by and large on various versions of one instrument, the Attitude Motivation Test Battery (AMTB), which aimed to assess individual differences in four classes which had been found to have motivational implications in SLA (Gardner, 2019). As outlined in Table 3.2, these four classes were integrativeness (three scales), attitudes toward the learning situation (two scales), language anxiety (two scales) and motivation (three scales). Using these scales, the AMTB consisted of a multicomponential questionnaire with over 130 items (Dörnyei & Ryan, 2015), including multi-item Likert, multiple choice and semantic differential scales (Ushioda & Dörnyei, 2012). These indices were then analysed to determine the relationship with other independent or dependent variables such as language aptitude or L2 proficiency (Ushioda & Dörnyei, 2012).

Table 3.2

Scales Contributing to the Affective Constructs in the Socio-Educational Model

Class of complex variables	
Integrativeness	Integrative Orientation
	Attitudes toward the Language Community
	Interest in Foreign Languages
Attitudes toward the Learning Situation	Teacher Evaluation
	Course Evaluation
Language Anxiety	Language Class Anxiety
	Language Use Anxiety
Motivation	Motivational Intensity
	Desire to Learn the Language
	Attitudes toward Learning the Language

As noted above, motivation in the AMTB is divided into three subcategories: Motivational intensity, Desire to learn, and Attitudes toward learning the L2. Ortega

(2014, pp. 169-170), for example, highlights the following positive and negative items in each of these three scales in the ATMB:

Motivational intensity:	“I keep up to date with French by working on it almost every day”
	“I don’t pay much attention to the feedback I receive in my French class”
Desire to learn:	“I want to learn French so well that it will become second nature to me”
	“To be honest, I really have little desire to learn French”
Attitudes towards learning the L2:	“I love learning French”
	“I hate French”

Students may thus be understood to be motivated due to the fact that they work hard to improve their level, because they are driven to learn it or because it appeals to them, or unmotivated due to a lack of effort to learn the TL, little desire to learn it, and a negative attitude towards it.

While it was originally developed for use with English speakers learning French as an L2 in a Canadian context, the test has subsequently been translated and adapted to a number of other contexts, including Japan, Brazil, Croatia, Poland, Romania, and Spain (Gardner, 2019). It has been praised for its rigor and systematicity in data-gathering and analysis, comparability and replicability of data, and generalizability to wider populations (Ushioda & Dörnyei, 2012). In addition, despite potential issues in validity, an important consideration in such tests using a self-rating method given the potential for self-perceptions to be inaccurate (Brown, 2007), the ATMB has been shown to have good psychometric properties, including construct validity (Dörnyei, 2005). As a result, the AMTB, alongside Gardner’s advanced quantitative data processing techniques, has set high research standards in the field of L2 motivation (Dörnyei & Ryan, 2015).

Though Gardner’s model of motivation and SLA largely stole the limelight in the social-psychological period, other important frameworks were Schumann’s (1978, 1986) acculturation model, Clement’s (1980) social context model, and Giles and Byrne’s (1982) intergroup model (Al-Hoorie, 2017). Firstly, Schumann’s (1978, p. 379) acculturation model predicts that learners will “acquire the target language to the degree they acculturate to the target language group”. The theory highlights two types of acculturation:

Type 1, whereby the individual is socially integrated in the other community and also psychologically open to acquiring the other language, and Type 2, whereby the other community is seen as a reference group whose attributes may be either consciously or unconsciously adopted (Gardner, 1985). The model provides a taxonomy of factors which control social distance and determine the extent to which the learner will acculturate, including dominance/subordination, integration pattern, degree of enclosure of both groups, degree of cohesiveness of SL learning group, size of SL learning group, degree of congruence of the two cultures and inter-group attitudinal evaluations (Zaker, 2016). Gardner (1985, p. 137) has consequently posited that the theory is more of a model of “language non-acquisition”, given that it describes the factors which inhibit language acquisition. Clement’s (1980) social context model put forward the concept of linguistic self-confidence, which maintained that being involved in a multilingual community may be a major motivational factor for those who come into contact with the L2 community (Guerrero, 2015). Specifically, the quality and quantity of contact between members was found to determine future desire for intercultural communication and the extent to which the speaker identified with the L2 group (Dörnyei & Ushioda, 2013). Thus, an individual who has frequent, high-quality interactions will have higher linguistic self-confidence, which will in turn increase their desire to learn the TL. Finally, Giles and Byrne’s (1982) intergroup model concerns members of minority ethnic groups in a multicultural environment and how successfully they acquire the dominant language. As is explained by Dörnyei and Ushioda (2013), the model states that when members exhibit weak in-group identification, ethnolinguistic vitality and boundaries, they are more likely to assimilate to the dominant culture and acquire more target-like features of the language. On the other hand, when these factors are stronger, that is, greater in-group identification, ethnolinguistic vitality and boundaries, members are more likely to adopt a code which deviates from the standard in terms of accent and grammar.

Despite its importance in motivation research, the social-psychological tradition began to receive criticism towards the end of the century, in particular in three widely discussed reviews by Crookes and Schmidt (1991), Dörnyei (1994a) and Oxford and Shearin (1994). These researchers criticised Gardner’s model as being too restrictive and detached from wider developments in psychology (Ortega, 2014). In addition, as highlighted by Crookes and Schmidt (1991, p. 470), research to date lacked validity in the sense that it was “not well-grounded in the real-world domain of the SL classroom”. Thus, although the approach highlighted the need to promote positive attitudes to the TL

culture, it offered few real insights for teachers (Ushioda & Dörnyei, 2012). These shortcomings paved the way for a new motivation research agenda to be reopened in the 1990s.

3.2.2. The Cognitive-Situated Period

Following the criticisms of the socio-educational model, the 1990s saw a turn towards what Dörnyei (2005) has called the cognitive-situated period. Of particular importance in this period were two interrelated trends, namely, the necessity of aligning current L2 motivational research with cognitive theories in mainstream motivational psychology and moving towards a more situated analysis of motivation in specific learning settings (Ushioda & Dörnyei, 2012). That is not to say that social-psychological perspectives were discarded altogether, but rather that research expanded to include the microcontext of the classroom and the cognitive processes involved in language learning (Al-Hoorie, 2017). Noteworthy contributions to motivation research in the cognitive-situated period include Dörnyei's (1994b) three-level framework of L2 motivation; Williams and Burden's (1997) social constructivist model; Noels et al.'s (2000) model based on self-determination theory; Ushioda's (1996, 1998, 2001) work on attribution theory; Autonomy theory; and Task motivation. Each of these contributions will now be discussed in turn.

Firstly, Dörnyei (1994b) conceptualised L2 motivation in terms of three distinct levels:

1. Language level: aspects such as language, culture and community, as well as Gardner's instrumental and integrative subsystems.
2. Learner level: individual characteristics and cognitive processes.
3. Learning situation level: course-, teacher- and group-specific motives rooted in the classroom setting.

While the former two levels drew heavily on previous theories by Gardner and Clément, the third level presents a detailed dimension based largely on research in educational psychology (Ushioda & Dörnyei, 2013). In this framework, motivational levels are separated given that they are each understood to play a vital role in one's overall motivation independently. For example, even in the same learning situation, a learner's motivation may be drastically affected by a change at another level, such as a change in the TL at hand. Thus, all things being equal at the language level and learner level,

differences at the learning situation level, such as having a good or bad teacher, may either boost or decrease the individual's overall motivation. The presentation of the framework prompted a collection of articles referred to as "The Modern Language Journal debate" (Dörnyei, 1994a, 1994b; Gardner & Tremblay, 1994a, 1994b; Oxford, 1994; Oxford & Shearin, 1994), which spanned a wide range of issues concerning the new movement and which became an important representation of the research in the field at that time (Dörnyei, 2019).

Secondly, according to Williams and Burden (1997), a constructivist view of motivation maintains that each individual is motivated in a different way, but that this motivation is also subject to social and contextual factors. Their social constructivist model of L2 motivation, which has been referred to as an example of the "paradigm-seeking spirit" of the decade (Dörnyei & Ushioda, 2013, p. 53), consisted of a framework of motivational factors organised in terms of internal factors (e.g., intrinsic interest of activity, mastery, self-concept, etc.) and external factors (e.g., significant others, the nature of interaction with significant others, the learning environment, etc.). Thus, while each individual will differ in their internal motivational factors, their actions are always carried out within the cultural and social milieu of the external factors (Brown, 2007).

Thirdly, Deci and Ryan's (1985) self-determination theory (SDT) is a theory rooted in existential, humanistic and organismic philosophies which sees humans as having inherent psychological needs which need to be satisfied in order for them to develop and thrive (Noels et al., 2019). It is concerned primarily with how humans pursue three vital psychological needs: autonomy (the experience of being in control of one's own behaviour), relatedness (the experience of belonging or being connected to others) and competence (the experience of being capable or accomplished) (Dörnyei & Ryan, 2015). For example, according to Niemiec and Ryan (2009), autonomous students are those who willingly devote time and energy to their studies, competent students are those who feel they can meet the challenges of their schoolwork, and relatedness gives students the feeling that their teacher likes, respects and values them. SDT has been applied to L2 research due to the collaboration of L2 motivation researchers Richard Clément and Kimberly Noels with self-determination experts Robert Vallerand and Luc Pelletier (Ortega, 2014). This fruitful partnership led to the development of an instrument specifically intended to measure the various components of SDT in L2 motivation: the Language Learning Orientation Scale (LLOS; Noels, Pelletier, Clément & Vallerand, 2000). Covering intrinsic motivation, extrinsic motivation and amotivation, the LLOS

included seven subscales of items, namely amotivation, external regulation, introjected regulation, identified regulation, intrinsic motivation: knowledge, intrinsic motivation: accomplishment, and intrinsic motivation: stimulation (Dörnyei, 2005). As is pointed out by Ortega (2014), Noels and her colleagues have provided numerous studies offering robust empirical support for SDT in L2 motivation. The framework has also been commended for its classification and organisation of language learning orientation in terms of a self-determination continuum and for highlighting features of the social learning setting which potentially influence learners' intrinsic or extrinsic motivation (Dörnyei & Ushioda, 2013). However, while Noels (2001) has highlighted the benefit of the model in that it focuses on the human motives behind motivational states, she has also acknowledged a potential weakness in its lack of “theoretical apparatus to accommodate societal attitudes towards the L2 and its speakers, the influence of the sociocultural milieu, and issues of ethnic vitality and identification” (Ortega, 2014, p. 178).

Fourthly, attribution theory is based on the work of Bernard Weiner (1985). It assumes that individuals attempt to understand the causal determinants of their previous successes and failures and that different kinds of causal attributions affect behaviours in different ways (Dörnyei & Ushioda, 2013). Within the school context, these attributions most commonly include ability, effort, task difficulty, luck, mood, family background and help or hindrance from others (Graham, 1994). As Dörnyei and Ushioda (2013) point out, while ability and effort have been found to be particularly dominant in western culture, variation has been noted in the stability and controllability of the attributions. For example, low ability is considered to be more stable and uncontrollable than effort, and consequently may hinder future achievement to a greater extent. Thus, students failing a test because they feel they are stupid will have a greater impact than if they fail because they have not studied hard enough. Within the context of L2 motivation, Ushioda's (1996, 1998, 2001) work with Irish learners of French has highlighted the important role of attributional processes in language learning. In particular, her work on the subject led to the identification of four attributional patterns in L2 learners (Ushioda, 2001, p. 120):

- a) enhancing one's self concept by attributing positive L2 outcomes and achievement to personal ability or qualities (e.g., hard work, effort)
- b) maintaining motivation by attributing negative L2 outcomes or lack of success to temporary shortcomings that might have changed (e.g., lack of effort, lack of opportunity to spend time in the L2 environment)

c) dissociating demotivating experiences by attributing negative experiences to the demerits of the institutionalised learning context (e.g., teaching methods, coursework pressures)

d) believing in a capacity for self-motivation through personal resourcefulness and initiative when confronted with demotivating experiences of institutionalised learning (e.g., setting oneself goals, engaging in intrinsically motivating TL activity)

In essence, these patterns illustrated ways in which learners take control of their affective learning experience, allowing them to capitalise on their success by attributing it to personal factors while at the same time minimising motivational damage when encountering setbacks in their learning by attributing them to shortcomings which can be overcome. According to Dörnyei and Ryan (2015), attribution theory has led the development of L2 motivation in two important directions. Firstly, it highlighted the temporal nature of motivation, linking learners' past experiences with their future achievement by means of causal attributions. Secondly, the limitations of using a survey methodology to assess causal explanations further highlighted the need for a qualitative approach in L2 motivation, a line of inquiry which would become increasingly important in subsequent approaches.

Fifthly, autonomy, as discussed above with regards to SDT, has also taken centre-stage in autonomy theory, which draws on the idea that “people who take the initiative in learning (proactive learners) learn more things and learn better than do people who sit at the feet of teachers, passively waiting to be taught (reactive learners)” (Knowles, 1975, p. 14). In other words, the success of a learner is dependent on their active engagement and involvement in the learning process. Drawing on cognitive theories and motivation concepts in the field of education, Dickinson (1995) set out to analyse the connection with autonomy and to justify its promotion in language learning (Dörnyei & Ushioda, 2013). The author reported substantial evidence from cognitive motivational research that learning success was conditional on learner autonomy. In particular, aspects such as taking responsibility for and being able to control one's own learning and perceiving that one's learning success or failures could be attributed to one's own efforts were found to be characteristic of learner autonomy in applied linguistics. Another important contribution to autonomy theory in this period was Littlewood's (1996) framework of autonomy, which is concerned with two components at the core of the notion of autonomy

(learners' ability and willingness to make choices independently) and three domains in which learners make and carry out choices (communication, learning and their personal life). These components and domains of autonomy provide a framework which aims to coordinate strategies for helping learners to develop autonomy. Concerning the components in the framework, while *ability* includes knowledge and skills, *willingness* includes motivation and confidence. Regarding the domains, *autonomy as a communicator* is dependent on the ability to use the TL creatively and to use appropriate strategies for communicating meanings in specific situations; *autonomy as a learner* is dependent on the ability to engage in independent work and to use appropriate learning strategies; and *autonomy as a person* is dependent on the ability to express personal meanings and to create personal learning contexts (Littlewood, 1996). The framework also indicates a high degree of interdependence between the different domains. For example, as learners develop more autonomy as communicators, they can also take better advantage of learning opportunities. As Dörnyei and Ushioda (2013) point out, the concept of autonomy has continued to gain ground in the field of L2 motivation and has evolved through the process-oriented period and socio-dynamic period of motivation, as will be discussed below.

Finally, in addition to the developments concerning theoretical perspectives mentioned above, such as attribution theory, SDT and autonomy theory, the cognitive-situated period also came with innovations to the learning situation, for example in the case of task motivation (Dörnyei & Ushioda, 2013). As an alternative to the presentation-practice-production approach, whereby a target feature is explicitly presented, practised in controlled production activities and then produced in free-production situational grammar activities, task-based language teaching seeks to develop learners' communicative competence by having them engage in meaning-focused activities via tasks (Ellis & Shintani, 2013). Thus, in a task-based approach the students essentially learn by doing; in other words, by engaging in meaning, the learner's system is encouraged to develop (Skehan, 1996). According to Dörnyei (2002), task motivation may be seen as the culmination of the situation-specific perspective that emerged in this period of motivation research, given the suitability of a task-based framework in examining motivation in a situated way. One benefit of the approach is that flow-introducing activities which involve task engagement and discussions have been shown to be beneficial for the learner's Willingness to Communicate (WTC), resulting in more meaningful communication when using the L2 (Yashima, 2019). Dörnyei and Kormos

(2000), in particular, have examined the ways in which tasks can be analysed in terms of motivational and other socio-dynamic variables (Cook & Singleton, 2014). In their research on task motivation, disposition toward task performance was found to have at least three different layers: generalised motives (integrativeness), course-specific motives (the appraisal of the L2 course) and task-specific motives (attitudes toward the particular task) (Dörnyei, 2001). In addition, as opposed to Julkunen's (2001) distinction between *trait motivation* (general motivational orientation) and *state motivation* (situation-specific motivation), which suggested a rather static conception of motivation, Dörnyei (2002) proposed a more complex view of task motivation (Dörnyei & Ushioda, 2013). He argued that, given that instructional tasks can last for a considerable amount of time, it is unlikely that an individual's motivation will remain constant throughout this period (Dörnyei, 2002). Instead, he advocated using an approach which looked into the dynamic motivational processes which are ongoing during task completion. It is this new shift towards the dynamic processes of motivation that brings L2 motivation into the next period of research: the process-oriented period.

3.2.3. The Process-Oriented Period

The turn of the century in L2 motivation research is characterised by a focus on motivational change, with efforts being made to capture the dynamics of L2 motivation either at the micro level, as in the case of task motivation, or the macro level, when analysing motivation during a course of study or across an individual's learning history (Dörnyei & Ushioda, 2013). As mentioned above, some research in the cognitive situated period had already begun to pave the way for these more process-oriented approaches. In addition, much of the research that was characteristic of this period may now be viewed as an early manifestation of, or precursor to, the current socio-dynamic period, discussed below. The methodology in this period also underwent a shift towards more suitable approaches, such as using quantitative instruments with a longitudinal approach to gather measures of attitudes and motivation at different points in time and using biographical or autobiographical methods or qualitative interviews (Dörnyei & Ushioda, 2013). According to Guerrero (2015), three notable perspectives during this period include Williams and Burden's (1997) focus on time, Ushioda's (1994, 1996a, 1998) model which also focused on time and Dörnyei and Ottó's (1998) process model of L2 motivation; each of which will now be discussed in turn.

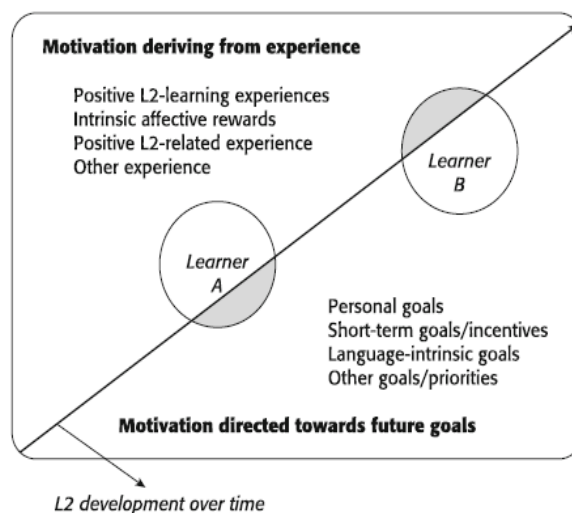
Firstly, in addition to their work on the social constructivist model (see [Section](#)

3.2.2), Williams and Burden (1997) were among the first to make the simple distinction between motivation *for* engagement (the reasons and choices for learning) and motivation *during* engagement (how the individual feels while learning) (Dörnyei & Ushioda, 2013). In their work on *focus on time*, the researchers placed notable importance on the temporal dimension, acknowledging the lengthy process of language learning and dividing the motivational process into three stages along a continuum: the reason for learning, the decision to learn the L2 and the effort that is sustained over time (Woodrow, 2017). In simpler terms, the authors explain that the model may be viewed as having two stages, namely initiating motivation and sustaining motivation (Williams & Burden, 1997). Emphasis is thus placed on the fact that motivation is not simply a matter of arousing interest, but rather also involves sustaining it by means of time and energy.

Secondly, alongside her work on attribution theory (see Section 3.2.2), Ushioda (1994, 1996a, 1998) also focused on the temporal motivation of the learner. This work culminated in a theoretical framework of motivation from a temporal perspective (Ushioda, 1998, p. 82), as outlined in Figure 3.4.

Figure 3.4

Ushioda's (1998) Theoretical Framework of Motivation from a Temporal Perspective



Note. From *Teaching and researching: Motivation* (2nd ed. p. 63) by Z. Dörnyei & E. Ushioda, 2013, Routledge. Copyright 2011 by Taylor & Francis.

As can be seen, the framework emphasised L2 motivation deriving from both the learners' experiences and from the motivation directed towards their future goals, each of which may take a more or less central role throughout time as the L2 develops. For example, in

the case of Learner A, motivation comes largely from positive experiences while goal-directed motivation plays a minor role. In the case of Learner B, the opposite is true, with goal-directed patterns taking a more prominent role. Notably, Learner B may represent a later stage of Learner A's thinking, as future goals may become increasingly important as the L2 develops (Ushioda, 2001). In addition to this temporal framework, Ushioda (1996) at this time had also begun to call for more introspective research approaches, highlighting the need to explore qualitative developments in the learner's dynamic motivational experience as it progressed over time.

Finally, perhaps the most illustrative work of the process-oriented period has been Dörnyei and Ottó's (1998) process model of L2 motivation, which again viewed motivation as a highly dynamic phenomenon (Cook & Singleton, 2014). The model contains two dimensions, the Action Sequence (which draws on Heckhausen and Kuhl's 1985 Action Control Theory) and Motivational Influences, which are presented along three key stages (Woodrow, 2017, p. 239):

1. The pre-actional phase, which reflects the activation of motivation.
2. The actional phase, which reflects executive motivational influences.
3. The post-actional phase, which reflects evaluation of the action.

These three phases have also been referred to as "choice motivation," "executive motivation," and "motivational retrospection" in later developments of the model (Dörnyei, 2002). The first phase includes goal setting, intention formation and initiation of intention enactment. It involves the learner's choice to start learning the TL and to set goals in order to achieve this. The second phase involves maintaining one's motivation, which is accomplished by generating and implementing subtasks, appraisal of one's achievements and action control or self-regulation. Finally, the third phase involves post-actional evaluation, whereby the learner reflects and evaluates their learning experience, either once a goal has been attained or the action has been terminated. It includes forming causal attributions, dismissing intention, and further planning and elaborating standards and setting strategies. According to Dörnyei (2005), his process model of L2 motivation has two main shortcomings. Firstly, it assumes that it is possible to clearly define when a learning process begins and ends. While this may be possible under controlled laboratory settings, it is by no means plausible in real classroom settings, where multiple learning processes may simultaneously run and where it is impossible to pinpoint the exact start and end of these processes. Secondly, it assumes that the actional phase occurs without

interferences from other actional processes which may be engaged at the same time. For example, these may relate to other academic pursuits, or other personal and social goals. Thus, the process model of L2 motivation could not do justice to the multifaceted dynamic and situated complexity of the ongoing processes which shape the learner's behaviour (Dörnyei & Ushioda, 2013). However, one key contribution of the model is the view of motivation in chronological stages, which highlighted the new understanding that motivation was not static. Instead, importance was placed on the development of self-regulatory strategies which could be used to support one's motivation throughout the learning process (Dörnyei & Ushioda, 2013). In addition, Guerrero (2015) highlights that this view of motivation from other perspectives at the learner's personal level raised awareness of the specific learning context of the learner, which would lead to the beginnings of the most recent period of L2 motivational research: the socio-dynamic period.

3.2.4. The Socio-Dynamic Period

As indicated above, earlier approaches to L2 motivation saw motivation and L2 achievement as having a cause-and-effect relationship, whereby a particular variable would influence the learner's motivation and then in turn their L2 achievement. However, such a stance evidently fails to capture the dynamic nature of motivation, in particular regarding contextual influences such as the setting and time (Dörnyei & Ushioda, 2013). In response to this issue, the socio-dynamic period reflects the current view of L2 motivation, where the central tenet is "a move from a linear view of motivation to a more complex set of interrelated learning and contextual variables" (Woodrow, 2017, p. 239). From the onset, a dynamic turn in SLA research was welcomed by researchers, as it had the potential to resolve numerous puzzling language learning issues (Dörnyei, MacIntyre & Henry, 2014). For example, it could disentangle the mystery of why L2 teaching input sometimes made a remarkable difference and other times made minimal or no difference at all. In addition, it offered a holistic approach which enabled researchers to unite the analysis of multiple elements or conditions relevant to specific situations, as opposed to examining well-defined variables to see whether there was a relationship between them, as had been done in the past. Dörnyei and Ushioda (2011) have highlighted three new approaches which differ significantly from earlier models and frameworks and which, in their view, centrally define this transition into a socio-dynamic period: Ushioda's (2009) person-in-context relational view of motivation, Dörnyei's (2009a) motivation from a

complex dynamic systems perspective and Dörnyei's (2005, 2009b) L2 Motivational Self System (L2MSS). Each of these approaches will be discussed below, with a particular focus on the L2MSS, given its central role in the current work.

Firstly, Ushioda's (2009, p. 218) person-in-context relational view of L2 motivation presents a "mutually constitutive relationship between persons and the contexts in which they act- a relationship that is dynamic, complex and non-linear". Thus, rather than viewing motivation as an innate individual characteristic, focus is placed instead on the individuality of each learner within the specific context. In addition, the context itself is no longer viewed as a static variable, but rather as a developing process in which the learner is involved in shaping through how they interact with the input (Ushioda, 2014). However, Dörnyei and Ryan (2015, p. 86) have questioned whether "a line of inquiry in which the agent cannot be meaningfully separated from the social environment within which he/she operates [can] be accommodated within existing paradigms". Thus, as is pointed out by Woodrow (2017), carrying out research which investigates this relationship is challenging, leading to little empirical evidence to date.

Secondly, a complex dynamic system (CDS) has been defined as "any system consisting of many interacting components" and in which "the interaction between the components changes their properties and creates properties on a macroscopic level" (van Geert, 2011, p. 274). As Ellis (2007) has remarked, a dynamic systems theory characterisation of second language acquisition pinpoints the coming of age in SLA research. Within this field, de Bot, Lowie and Verspoor (2007) explain that a CDS may be applied to language as follows: language learners are understood to have their own cognitive ecosystems made up of intentionality, cognition, intelligence, motivation, aptitude, the L1 and the L2. This ecosystem is in turn related to factors such as language exposure, maturity and level of education, which are again in turn related to the learners' social ecosystem which consists of the environment where learning takes place and with which the learner interacts. The approach thus accounts for the interconnectivity of language learning factors such as motivation and language input, while also accounting for the dynamic interplay of these multiple factors. This idea was taken up by Dörnyei (2009a) in his treatment of motivation from a CDS perspective, which maintains that there are three broad distinctions between motivation, cognition and affect. Each of the three is seen as a dynamic subsystem, which is never in isolation but rather is in constant interaction with the other subsystems (Dörnyei, 2010). In order to capture this complexity, Dörnyei (2009a, p. 225) proposes the identification of "higher level amalgams or

constellations of cognition, affect and motivation that are relatively stable and which act as ‘wholes’”. Given the dynamic complexity of the ongoing mental processes and attribute, he argues that we are unable to meaningfully distinguish more than these three key dimensions (Dörnyei & Ushioda, 2013). This CDS approach to L2 motivation has provided the necessary radical change to overcome issues with previous approaches, which had viewed motivation as “a patchwork of interwoven cause-effect relationships” (Dörnyei, 2009a, p. 210). However, while in theory the approach offered a more realistic representation of the complex ongoing processes, in practice this was more difficult to capture. Thus, despite ample coverage of CDS in SLA literature, by the end of the 2010s there was a sense that very little of this work was empirically tested. As Dörnyei et al. (2014, p. 1) remark, “scholars spent much more time *talking* about research in a dynamic systems vein than actually *doing* it”. Thus, in their 2014 volume *Motivational Dynamics in Language Learning*, Dörnyei, MacIntyre and Henry aimed to tackle precisely this problem, by drawing together the work of numerous scholars taking a truly dynamic approach to motivation and encouraging the uptake of this promising avenue in future research.

Finally, Dörnyei’s (2005) L2MSS, though rooted firmly in earlier L2 motivation research, presented a major reformation to previous research by explicitly using psychological theories of the self (Dörnyei, 2009b). In addition, it sought to overcome issues in previous theories regarding the concept of integrativeness. The trigger for this was the results of a large-scale motivation survey in Hungary, carried out with over 13,000 students over a period of 12 years (Dörnyei et al., 2006). Despite the dissimilarity between the monolingual Hungarian context in Dörnyei’s work and the Canadian context in Gardner’s motivation research, results concerning the notion of integrativeness were very similar in both cases (Csizér, 2019). This led Dörnyei and Csizér (2002) to posit an alternative, broader sense of integrativeness which could explain these findings. This new system is influenced in particular by two important psychological concepts: Markus and Nurius’ (1986) theory of possible selves and Higgins’ (1987) self-discrepancy theory. The former theory of possible selves is concerned with the future self-conceptions an individual has, which are divided into three types: desired possible selves (selves which we would like to become), feared possible selves (selves which we are afraid of becoming) and selves that we could become (Hessel, 2015). These possible selves are understood to directly influence motivation and behaviour, thus, if one’s desired possible self includes “the rich self”, this desired self can motivate the individual to reach the goal of becoming

rich. The latter self-discrepancy theory distinguishes between the *ideal self* (the attributes we would ideally like to possess), the *ought self* (the attributes we feel we ought to possess) and the *actual self* (the way we currently see ourselves). While the *ideal self* is concerned with the individual's own vision for himself or herself, the *ought self* involves someone else's vision. According to the theory, discrepancies between the various constructs is undesirable and will thus lead to the desire to minimise any discrepancies between the actual and ideal or ought to self. Dörnyei's L2MSS combines aspects of these two theories to form the idea that the language learner's motivation consists of three key parts:

1. The Ideal L2 Self (the image we have of our future self as an L2 user based on our own wishes)
2. The Ought-to L2 self (the image we have of our future self as an L2 user based on external expectations)
3. The L2 Learning Experience (concerning 'executive' motives such as the effect of the language teacher, curriculum, peer group, experience of success or failure)

Firstly, the Ideal L2 Self typically fosters integrative and internalised motives. For example, individuals may be motivated to learn the TL because they can visualise themselves being surrounded by lots of friends who speak the TL, or because the TL really appeals to them. Secondly, the Ought-to L2 Self reflects more extrinsic types of instrumental motivation, dealing with attributes which individuals feel they must possess so as to meet expectations and avoid negative outcomes. For example, individuals may feel like they need to work hard at learning the TL so that they do not disappoint their parents or because there is a societal expectation that they should learn the TL. Finally, the L2 Learning Experience is concerned with contextual issues. For example, individuals may not want to learn the TL because their teacher is not very nice or because they always get bad grades. As Csizér (2019) points out, given that the model includes not only the first two self-related constructs but also this third one related to the learning environment, it acknowledges the fact that the two self-guides do not affect the learning process in isolation, but rather highlights the interaction between the different constructs. Following its elaboration, the model was implemented in numerous studies which sought to provide empirical evidence for the central constructs. Ryan (2008) was the first to develop scales designed to measure the Ideal and Ought-to Self in a Motivational Factors Questionnaire

(MFQ). For example, the scale measuring the Ideal L2 Self, which had a Cronbach Alpha of .80, included the following six statements:

The things I want to do in the future require me to speak English.

Whenever I think of my future career, I imagine myself being able to use English.

I often imagine myself as someone who is able to speak English.

If my dreams come true, I will use English effectively in the future.

I can imagine speaking English with international friends.

When I think about my future, it is important that I use English.

According to Dörnyei and Ushioda (2011), several studies have been carried out specifically to test the L2MMS in an array of contexts (e.g., Csizér & Kormos, 2009; MacIntyre, MacKinnon & Clément, 2009; Ryan, 2009; Taguchi, Magid & Papi, 2009) and have largely attested its validity, finding Integrativeness and the Ideal L2 Self to be closely related, and that the Ideal L2 Self consistently explained criterion measures better than Integrativeness. In addition, while in all studies the Ideal L2 Self had a higher correlation with Instrumentality-promotion than with Instrumentality-prevention, the opposite was found in the Ought-to L2 Self.

The L2MSS, as well as the research set out in Dörnyei and Ushioda's 2009 volume *Motivation, Language Identity and the L2 Self* showcasing its validity, set about a surge in motivation research over the next decade, which was largely dominated by the L2MSS model (Boo et al., 2015). As Csizér (2019) noted, the items used in the MFQ have been continuously revised and revalidated for each new context where research is carried out, which have been discussed and summarised by Papi, Bondarenko, Mansouri, Feng and Jiang (2019). One such example in a Spanish context is research by Brady (2014, 2015, 2019), which involved the design and validation of an MFQ used to investigate the L2MSS in over 500 Spanish learners of English. Based on Ryan's (2009) work in Japan and Taguchi et al.'s (2009) work in Asia, the MFQ presents 67 items across 13 scales, which target the Ideal L2 Self, the Ought-to L2 Self, and a range of goal-related and affective motivational variables.

Despite the empirical validity of the L2MSS, and the enthusiasm for its implementation, some issues have also been raised, which Csizér (2019) groups into

conceptual and measurement ones. Firstly, some researchers maintain that the model is missing some self-related components. For example, MacIntyre, MacKinnon and Clément (2009) argue that a wider range of possible selves should be recognised, rather than focusing solely on the ideal and ought-to selves. In addition, issues have repeatedly been reported regarding the Ought-to Self, with some studies often omitting the component due to unacceptable Cronbach Alpha values (e.g., Kormos & Csizér, 2008; Csizér & Lukács, 2010). This issue, when occurring among secondary school participants, is suggested to occur due to the fact that learners are potentially too young to internalise the pressure which the environment places on them (Csizér & Lukács, 2010). Secondly, regarding measurement issues, Al-Hoorie (2018) highlights the problematic approach of using intended effort as a criterion variable, especially when intended effort is measured by self-report. In such a case, it is simply assumed, rather than demonstrated, that L2 achievement is related to the learner's intended learning efforts (Moskovsky, Assulaimani, Racheva & Harkins, 2016). It is instead necessary to incorporate some kind of language proficiency test alongside the MFQ, as is done by Lamb (2012), who uses a C-test to determine the relationship between L2 proficiency and the components of the L2MSS. The benefit of this approach is that it provides a more accurate picture of the students' language ability, rather than focusing solely on their self-related reports of proficiency and intended learning effort. In addition, further issues have been raised regarding the prevalence of studies focusing on English, which evidently leads to the neglect of other TLs. This is evident in Al-Hoorie's (2018) meta-analyses of the L2MSS, which included only three studies investigating a language other than English, despite providing a very thorough analysis of research on the L2MSS. As highlighted by Ushioda and Dörnyei (2017), this is evidently problematic given that current theoretical perspectives, designed specifically for the analysis of English, may not be suitable for other languages. In addition, little is known regarding the effect that global English has on the motivation to learn other languages. This issue will be dealt with in greater detail in Section 3.4.2.

This section has offered a comprehensive overview of the history of L2 motivation. It began first with the initial stages of L2 motivation research in the socio-psychological period, highlighting the revolutionary work of Gardner and his colleagues in the Canadian context. It then turned to the cognitive-situated period, with its emphasis on the context and cognitive theories in mainstream motivational. It next highlighted the shift towards a more process-oriented approach at the turn of the century and finished with the subsequent socio-dynamic period that represents the current thinking on L2 motivation

research. Bearing in mind this historical and theoretical overview, the following four sections will turn to more empirical research on motivation. Section 3.3 will offer an introduction to L2 motivation research and measurement, highlighting the advantages and disadvantages of both quantitative and qualitative approaches to measuring L2 motivation. The final three sections will then focus on motivation research in relation to three issues which are of central importance to this work: motivation in English and languages other than English (LOTEs), the language learning context, and gender.

3.3. L2 Motivation Research

As outlined in Section 3.2, each new period in motivation research brought with it new instruments for measuring motivation, the most important being the AMTB, the LLOS, and the more recent MFQs investigating the components of the L2MSS. While the instruments in each new period evolved and reflected new understandings in L2 motivation, what they all have in common is their predominantly quantitative approach in using questionnaires and language tests (Lasagabaster, Doiz & Sierra, 2014). According to Dörnyei and Ushioda (2011), this reliance on self-report data in questionnaires when analysing motivation can largely be attributed to two factors. On the one hand, it is particularly challenging to measure motivation objectively, especially when considering more recent views of motivation as a multifaceted construct which is subject to a wide range of internal, contextual and temporal processes. On the other hand, given the variety of motivational factors at play, it is not directly observable. For example, a student who raises their hand to answer a question may do so due to an interest in learning, to show that they know the answer, to outperform their peers, or to please their teacher. The advantage of using a survey-based quantitative approach is thus that it allows us to target a wide range of language-related issues and in turn draw conclusions about larger L2 learning populations (Dörnyei & Csizér, 2012). As Fryer, Larson-Hall and Stewart (2018, p. 56) point out, as opposed to qualitative approaches, this can be done “not just with a handful of subjects anecdotally, but with a broader sample of the population after accounting for a variety of alternate reasons the phenomena could have occurred”. This can evidently provide ample data from which to make decisions and form policies in an informed and well-justified way (Dörnyei & Csizér, 2012). However, a main disadvantage of the approach is the inherent subjectivity and reliability of self-report data. According to Dörnyei and Ushioda (2011), this issue has been addressed via the construction of rigorous measurement instruments which have been widely tested and

validated, such as the AMTB. An additional shortcoming occurs in studies adopting a cross-sectional approach. Studies using this design provide only a snapshot of the learners' motivation, which fails to account for the dynamic and changing nature of motivation. In order to combat this issue, longitudinal studies provide data which can measure the temporal dimension of learners' motivation; an approach which is almost always preferable to a cross-sectional one in applied linguistics (Ortega & Iberri-Shea, 2005; Ortega & Byrnes, 2008; Fryer et al. 2018). As an alternative to this predominant quantitative perspective, and particularly given the flourish of process-oriented and socio-dynamic research, researchers have increasingly taken a more qualitative approach. In Boo et al.'s (2015) review of L2 motivation research, they highlight both the repeated calls for such investigation, as well as the consequential substantial increase over the period under scrutiny. The data collection process in this research is largely dominated by ethnographies and interviews (Fryer et al., 2018). What these approaches have in common is a view of language learning as a social process, and a focus on the context of learning (Friedman, 2012). An obvious advantage of such a perspective is that it allows us to gain a better understanding of the dynamic, ongoing motivational processes of the learner (Kim, 2009). This type of rich data can provide greater data-driven depth to the analysis, by analysing the participants' unique experiences (Dörnyei & Ushioda, 2013). However, qualitative methods have also been criticised for their subjectivity, which may be somewhat inevitable (Lew, Yang & Harklau, 2018). As Dörnyei and Ushioda (2013, p. 205) noted, "there are few firm safeguards inherent in the approach to make sure that results are not influenced by the researcher's personal biases and idiosyncrasies". In addition, the greater depth and volume of data in qualitative analyses is evidently much more time-consuming, which entails much smaller participant samples than what is possible in the numerical data found in quantitative studies (Friedman, 2012).

As can be seen, both a quantitative and qualitative approach come with advantages and disadvantages. However, as Fryer et al. (2018) have noted, rather than adopting a one-or-the-other approach, researchers such as Brown (2004) have posited a view of the approaches across a continuum, whereby each places value on different aspects. Thus, while quantitative research may prioritise generalisability, reliability and validity, qualitative research places greater emphasis on dependability, credibility and confirmability.

As is clear from Boo et al.'s (2015) review of L2 motivation research and Al-Hoorie's (2018) meta-analysis of the L2MSS, which cover 416 and 678 relevant sources

respectively, recent years have seen a drastic increase in L2 motivation research. The following three sections offer a review of some this research, focusing on the areas which are most pertinent to the current work: motivation in English and LOTEs, the learning context, and gender. For each of these three issues, a theoretical background will be provided in order to contextualise the topic in the field of L2 motivation, which is then followed by a review of the most relevant research.

3.4. Research on Motivation in English and LOTEs

The dominant status of English in the current era of globalization and multilingualism has had a significant impact on the field of L2 motivation (Ushioda & Dörnyei, 2017). The following two sections provide an overview of this impact, dealing first with the theoretical background of the learners' motivation in English and LOTEs and then turning to research that has investigated this issue. This research is of central importance to this thesis, given that it aims to compare the language motivation of learners studying English as an L2 and French as an L3.

3.4.1. Language Motivation in English and LOTEs

Dörnyei and Ushioda (2013) remark that the spread of English as an international lingua franca is likely to have at least two repercussions on language learning motivation. Firstly, there may well be a qualitative difference between the motivation to learn English as opposed to any other language, due to the fact that knowledge of English is increasingly seen as a fundamental educational skill which is essential for professional advancement. This idea has also been addressed by Block and Cameron (2002), who suggest that language learning and communication skills which become necessary because of globalisation will influence the learners' motivation towards instrumentality. This supposition has been corroborated in numerous studies to date (e.g., Taguchi et al., 2009; Mehrpour & Vojdani, 2012; Islam, Lamb & Chambers, 2013; Warshani Himanshi Pilimalawwe, 2015). Secondly, due to the use of English in non-native speaker groups, traditional motivational concepts like integrativeness and attitude towards the TL community become increasingly obscure. This is because, rather than having a desire to learn English so as to be able to communicate with people from a specific English-speaking country, learners may wish to do so in order to communicate to other non-native speakers of English from various TL communities.

More recently, turning the focus specifically towards LOTEs, Dörnyei and Al-

Hoorie (2017, p. 457) have highlighted five vital aspects in the conceptualization of motivation in LOTEs:

1. the interaction of English- and LOTE-related self-images,
2. the individualistic focus of the Ideal L2 Self,
3. the Ought-to Self in languages with substantial versus marginal social support,
4. the goals in the learning of English and LOTEs, and
5. the unconscious motives in the study of English and LOTEs.

Firstly, what is most unique about motivation to learn LOTEs is perhaps that fact that it almost always takes place, as the authors note, “in the shadow of global English” (Dörnyei & Al-Hoorie, 2017, p. 457). As Henry (2010) points out, when an L3 self-image is appraised referentially against an English-speaking ideal-self, it may consequentially be seen as inferior, a comparison which has been found to have a potential negative impact on FL motivation. Secondly, the ideal-self concept emerged primarily as a response to global English, given the unsuitability of the concept of integrativeness. However, when dealing with LOTEs, which are often associated with a specific L2 community or learned for more integrativeness reasons, this shift may be detrimental in our explanation of L2 motivation (Dörnyei & Al-Hoorie, 2017). In addition, as it stands, the Ideal L2 Self portrays a specific future self of one individual using one single L2, and thus largely limits research to a monolingual bias, rather than exploring the learners’ multilingual selves. More recently, Henry (2017) has put forward the concept of the “Ideal Multilingual Self” in an effort to address this gap. This Ideal Multilingual Self concerns the learner’s aspirations to be or become multilingual and maintains that learners can develop self-guides for each TL they are learning, for example, an Ideal L_x Self and an Ideal L_y Self. Thirdly, the Ought-to Self is also suggested to be different when comparing English and LOTEs. This is because while English conceives a generally stable Ought-to Self-image, given the relatively constant support around learning English, the Ought-to Self in LOTE learning may be much more fragmented. Dörnyei and Al-Hoorie (2017) consequently suggest that, rather than one ought-to L2 self in LOTEs, it may make more sense to speak of ought-to *selves*, acknowledging the multiple, potentially conflicting ought-to L2 self-images that may be at play. This notion has already been addressed by Dörnyei and Chan (2013, p. 456), who note that

in light of the ambiguities that have surfaced with regard to the ought-to self, it would have been better to apply more elaborate scales targeting different types of external pressures separately ... instead of using a single ought-to self scale.

Fourthly, global English is often considered the default option in language learning. For example, it is commonplace for an individual to have studied English but not another language, but the reverse is something of an anomaly. As a result, those individuals who do actively obtain a high level of LOTE proficiency generally do so for highly specific and personalised reasons (Dörnyei & Al-Hoorie, 2017). As a result, and related to the discussion on the ought-to LOTE self above, it may be necessary to adopt more finely tuned and specific instruments when investigating LOTE motivation. Finally, as has been demonstrated by Zajonc (1968), the *mere exposure effect* entails that being exposed to a stimulus repeatedly increases favourable attitudes toward that stimulus, whether it is supraliminal or subliminal (Zajonc, 2001). Dörnyei and Al-Hoorie (2017) consequently note that, given the pervasive presence of global English, there are likely marked differences in the unconscious domains of learning English as opposed to LOTES, which should again be taken into consideration when offering an account of language motivation in different languages.

3.4.2. Language Motivation Research on English and LOTES

Regarding language motivation research, the dominance of English has again taken centre stage. On the one hand, its status has largely influenced current conceptualisations of L2 motivation in relation to the concepts of self and identity. On the other hand, its importance has consequently entailed the neglect of LOTES, which remain “a largely uncharted area of language learning motivation” (Ushioda & Dörnyei, 2017, p. 451). In an effort to address this imbalance, *The Modern Language Journal* recently released a special issue concerning the motivation to learn LOTES in an era of globalization and multilingualism. In particular, the issue aimed to tackle two key questions, namely whether current theoretical perspectives were suitable for analysing motivation in LOTES, and what effect global English has on the motivation to learn other languages. This issue, and its calls for further investigation of this subject, has been key in prompting the necessary research on LOTE motivation, much of which has sprung up since its publication. This recent research addressing language motivation in LOTES has either attempted to analyse the L2 motivation of LOTES in isolation or has studied it comparatively alongside English. The following two sections will offer an overview of

this research, focusing first on research which deals with motivation towards LOTEs in isolation, and then on that which compares motivation towards English and LOTEs.

3.4.2.1. Research on Motivation towards LOTEs. This section addresses the limited research that has been carried out on LOTEs in isolation, without any direct comparison with the L2 (English). Some of this research is summarised in Table 3.3 and discussed below.

Table 3.3

Research on Motivation towards LOTEs

Authors	Languages	Focus	Main Findings
McEown & Sugita-McEown (2020)	L1: Japanese L2: English L3: Chinese / Spanish / French / German	Positive and negative psychological factors on effort and FL anxiety in LOTES	Intrinsic motivation endorsed to a greater extent than other emotional orientations; teacher support positively predicted intrinsic motivation for all LOTES.
Wang & Liu (2020)	L1: Chinese L2: English L3: German	Ideal and Ought-to L3 selves and their connection with the L3 Learning Experience.	Fluctuation of the learners' L3 motivation over the two-year period: learners' ideal L3 selves initially showed a clear upwards trajectory, and then later declined; ought-to L3 selves also became weaker over the two-year period in response to their L3 Learning Experience.
Zhang (2020)	L1: Chinese L2: English L3: Italian/Thai	L3 motivation using the L2MSS and self-discrepancy model	L3 Learning Experiences shaped learners' ideal and ought-to selves; the four facets of the self-discrepancy model may co-exist.

Firstly, McEown and Sugita-McEown (2020) investigated the effect of positive and negative psychological factors on effort and FL anxiety in LOTEs in 221 Japanese learners. In addition to studying English as an L2, participants were also studying either Chinese ($n = 41$), Spanish ($n = 76$), French ($n = 62$) or German ($n = 42$) as an L3. Participants completed a questionnaire which included the constructs integrative orientation, intrinsic motivation, introjected regulation, teacher and parental support, intended effort for learning the LOTE, and LOTE anxiety. Results indicated that intrinsic motivation was endorsed to a greater extent than other emotional orientations such as integrative orientation and introjected regulation. In other words, most of the students learning LOTEs did so for their own enjoyment. This was especially so in the case of French and German learners. In addition, teacher support was found to positively predict intrinsic motivation for all LOTEs, while for parental support this was only the case for learners of Chinese and Spanish. Regarding FL anxiety, introjected regulation was a significant positive predictor only in the case of learners of Spanish.

Wang and Liu (2020) took a more qualitative approach to analysing L3 motivation, using semi-structured interviews to investigate the Ideal and Ought-to L3 selves and their connection with the L3 Learning Experience. Participants included five university students in a Chinese university who were studying German as an L3 and who took part in three rounds of interviews over two years. Results highlighted the fluctuation of the learners' L3 motivation during the period under investigation. While learners' Ideal L3 selves initially showed a clear upwards trajectory, a decline was observed later on. Ought-to L3 selves also became weaker over the two-year period, in response to their L3 Learning Experience. The authors highlight the negative influence of global English as a contributing factor, pointing out that while learners could see the benefit of being multilingual, the influence was not enough to persuade them to work harder at learning German.

Finally, Zhang (2020) also used interviews to investigate the L3 motivation of two Chinese students studying Italian and Thai. The participants first completed a questionnaire which adhered to Dörnyei's (2009) L2MSS framework. The results of this were then used to design the qualitative interviews, which were later analysed using Lanver's (2016) self-discrepancy model for language learners. The results indicated that the participants' L3 Learning Experiences shaped their ideal and ought-to selves, and highlighted the possible coexistence of the four facets of the self-discrepancy model (Own-Ideal, Other-Ideal, Own-Ought and Other-Ought). The authors thus highlight the

possibility of internalizing external expectations or pressure while maintaining strong self-directed motivation in L3 learning.

3.4.2.2. Research Comparing Motivation towards English and LOTEs. In addition to the above research which aims to analyse L3 motivation, other studies have carried out a comparison of the motivation towards the L2, generally English, as compared with the L3. Again, with the exception of researchers such as Alastair Henry, whose doctoral work over a decade ago began to deal specifically with this under-researched topic, research on motivation towards English as compared with LOTEs is particularly sparse and has only really begun to appear in recent years. A summary of some of this research addressing L2/L3 motivation, in particular that which compares the motivation of learners studying English as well as one other TL, is summarised in Table 3.4 and then discussed below.

Table 3.4

Research Comparing Motivation towards English and LOTEs

Authors	Languages	Focus	Main Findings
Henry (2010)	L1: Swedish L2: English L3: German or Spanish	Effect of global English on the L3	Significant inverse relationship between language learning motivation and negative appraisals of FL selves in relation to the normative function of English.
Henry (2011)	L1: Swedish L2: English L3: Spanish, French or Russian	Ongoing processes in L3 learning and the working self-concept	English plays an active role in the working self-concept in L3 learning; maintaining a viable L3-speaking/using self sometimes became difficult due to the referential effect of English.
Henry (2014a)	L1: Swedish L2: English L3: German or Spanish	Crosslinguistic awareness and the negative impact of the L2 on the L3 self-concept	Learners were aware of making crosslinguistic comparisons; the effect of the L2 had potentially detrimental effects on L3 motivation, though some participants used strategies to counteract these effects.

Table 3.4 (continued)

Henry (2014b)	L1: Swedish L2: English L3: French	Complex dynamics of the learners' fluctuating motivation	Participants frequently compared their L2 and L3 and viewed English as more enjoyable and important. These perceptions affected the L3 ideal self, the L3 Ought-to Self and, in some cases, the L3 Learning Experience.
Thompson & Erdil-Moody (2016)	L1: Turkish L2: English L3: Varied	Motivation, multilingual status and perceived positive language interaction	Significant differences in the Ideal L2 Self when comparing the bilingual and multilingual groups, as well the group who perceived a positive interaction between the FLs studied and the group who did not. No significant differences were found for the ought-to-self.
Busse (2017)	L1: Bulgarian, German, Dutch or Spanish L2: English L3: Varied	Affective dispositions, ideal English selves and perceptions of external pressure	The majority of students' attitudes towards English differ from those towards LOTEs, with English often being perceived as more important than other languages.
Sugita McEown et al. (2017)	L1: Japanese L2: English L3: Varied	Identity-relevant dimensions in the L2 motivational process	English had a somewhat negative impact on LOTE motivational processes; the interplay of self- and identity-related orientations between English and the other TL was potentially negative or positive when the main TL was more politically powerful than the other TL.

Table 3.4 (continued)

De Smet et al. (2018)	L1: French L2: English L3: Dutch	Comparison of anxiety and enjoyment in two TLs (English and Dutch)	CLIL students reported significantly less anxiety than non-CLIL students overall, those in English CLIL reported significantly less anxiety and more enjoyment than students in Dutch CLIL.
De Smet et al. (2019)		Language attitudes and motivation in two TLs (English and Dutch)	More positive attitudes and higher motivation were reported by CLIL than non-CLIL students, and by those in English CLIL compared to Dutch CLIL; effect sizes indicated that the TL plays a more crucial role in language attitudes and motivation than the educational approach does.
Geoghegan (2018)	L1: Spanish + Catalan L2: English L3: French or German	Motivation and identity in L2 or L3 study abroad	Statistically significant difference between the two groups with regards to the Ideal L2 Self, with the English group reporting a higher overall mean.
Geoghegan & Pérez-Vidal (2019)		Motivation, identity and English as a lingua franca	English as a lingua franca, plays a number of roles, both positive and negative. Participants in non-English speaking countries reported that English could be used as a crutch, which helped them in times of L3 language-related difficulties.

Table 3.4 (continued)

Calafato & Tang (2019)	L1: Arabic L2: English L3: French, Spanish, German, Japanese, Dutch, Italian, Russian, Turkish or Korean	Motivational self-concept and gender	When comparing English and the students' L3, statistically significant correlations were found only between female participants' ought to-L3 and English selves.
Bui & Teng (2019)	L1: Cantonese L2: English L3: Japanese	Motivation and self-perceptions of the L2 and L3 selves	Motivational dynamics are particularly complex in multilingual scenarios, in particular with regards to factors such as perceived ease and language status.
Oakes & Howard (2019)	L1: Swedish or Polish L2: English or French	Applicability of the L2MSS in French	Statistically significant differences between in four of the seven constructs: desire for proficiency and weak integrative (French learners had a higher mean) and instrumental and ought-to L2 self (English learners had a higher mean).

Henry (2010, 2011, 2014a, 2014b) has carried out numerous studies which investigate the L2/L3 motivation of native Swedish students. While the first of these studies used a quantitative methodology, the latter three all adopted a more qualitative approach. Firstly, Henry (2010) assessed the impact of global English on the self-concepts of other subsequent TLs. Participants included 182 6th grade students who had studied

English for between six and nine years, and French, German or Spanish for four years. Three years later, 101 of these students were tested again in order to collect longitudinal data. Two instruments were used: the first was modelled on the AMTB, though it incorporated Dörnyei's Ideal L2 Self, and the second was the specifically designed L2 self-concept referential scale, which encompassed the three elements of the L2MSS and aimed to determine the presence and effects of an L2 working self-concept. Results indicated that there was a significant inverse relationship between language learning motivation and negative appraisals of FL selves in relation to the normative function of English. Results from the longitudinal data also demonstrated a significant decline in attitudes to the learning situation but highlighted no such change in conceptions of an ideal FL self. Henry (2011) provides a more qualitative approach to this issue, by conducting semi-structured interviews. The four participants in the study again had Swedish as an L1, English as an L2 and, in this study, Spanish, French or Russian as an L3. The focus of the study was to analyse the ongoing processes in L3 learning by using the working self-concept. The study provides rich data, which highlights the role of English as an active constituent in the working self-concept in L3 learning. In some cases, it was highlighted that maintaining a viable L3-speaking/using self became difficult due to the referential effect of English. Henry (2014a), also using a qualitative approach, investigated the effect of crosslinguistic awareness and the negative impact of the L2 on the L3 self-concept. Participants included 21 8th grade students who were studying either German or Spanish as an L3. Semi-structured interviews were carried out, which firstly dealt with the participants' attitudes to the L3, learning situations, opportunities for use and future aspirations, and secondly with their crosslinguistic awareness in productive and receptive skills. Results indicated that nearly all participants were aware of making crosslinguistic comparisons. In addition, although the effect of the L2 was found to have potentially detrimental effects on L3 motivation, a number of participants were seen to use strategies to counteract these effects. The author highlights the importance of raising awareness of these issues so as to enable students to develop and adopt counteracting strategies. Finally, Henry (2014b) investigated the motivation of six secondary school students, who were learning French as an L3 and as a compulsory part of their programme. Using a qualitative approach, five to six semi-structured interviews were carried out with each participant. By adopting a CDS perspective, the author could examine the complex dynamics of the learners' fluctuating motivation. For example, temporary shifts in motivation were observed as a result of numerous factors such as perceptions relating to

how enjoyable/meaningful/relevant a particular activity was, the impact of perturbing factors such as receiving test results, and the interaction with other complex systems such as comparisons with English. Regarding the latter, participants were found to frequently compare their L3 with their L2, viewing English both as more enjoyable and more important and in some cases even view learning French as pointless. These perceptions were found to affect the L3 Ideal Self, the L3 Ought-to Self and, in some cases, the L3 Learning Experience.

Thompson and Erdil-Moody (2016) investigated language learning motivation and multilingual status using Dörnyei's L2MSS. Participants included 159 Turkish learners of English as an L2, 96 of whom who were also studying German, French, Spanish, Arabic, Russian, Italian or Korean as an L3. The instrument used was a 20-item questionnaire, based on Dörnyei and Taguchi (2010), investigating the ideal and ought-to selves. Responses were analysed in order to investigate the learners' motivation, and two operationalisations of multilingualism: previous LOTE learning experiences and Perceived Positive Language Interaction (PPLI). To this end, participants were divided into different groups: bilingual ($n = 63$) and multilingual ($n = 93$), and participants who perceived a positive interaction between the FLs studied (PPLI, $n = 70$) and those who did not perceive a positive interaction between the FLs studied (NPPLI, $n = 89$). Regarding the differences between bilingual and multilingual groups, results indicated that there was a significant difference for the Ideal L2 Self, with the multilingual group reporting a stronger Ideal L2 Self, but not for the ought-to L2 self. Similarly, when comparing the PPLI and NPPLI groups, statistically significant differences were found in the Ideal L2 Self, to the advantage to the PPLI group, but not for the Ought-to Self.

Busse (2017) analysed the survey responses of 2,225 secondary school students from Bulgaria ($n = 403$), Germany ($n = 685$), the Netherlands ($n = 560$) and Spain ($n = 607$). In addition to analysing the students' affective dispositions toward English, their ideal English selves and their perceptions of external pressure, the survey also included open-ended questions asking participants how their responses would differ if they were answering with regards to another FL they are studying. Results indicated that 80% of participants stated that they would have responded differently if the survey had focused on a LOTE, suggesting that the majority of students feel their attitudes towards English differ from those towards LOTEs. This was particularly so in students from Bulgaria and the Netherlands. When explaining this stance, participants pointed to the importance of English (in some cases even highlighting that it is more important than their other FL),

its status as a world language, and needing English in order to meet job requirements. In Spain, specifically, responses were somewhat more mixed, with many participants indicating that learning English is important, but not necessarily stating that it was more important than learning other languages. However, in some cases, participants referred negatively to French, the most commonly studied second FL in Spain, stating that as French is given less importance in the world, they therefore give it less importance.

Sugita McEown, Sawaki and Harada's (2017) study focused on the identity-relevant dimensions in the L2 motivational process in both English and LOTEs. Participants included 250 undergraduate students at a private university in Tokyo, who completed a questionnaire with 14 subscales. In order to address both English and LOTEs, subscales for self- and identity-related orientations were included twice, for the two types of languages. Structural equation modelling was used to integrate three self- and identity-related frameworks to evaluate the participants' LOTE motivation. Results found that the status of English had a somewhat negative impact of the learner's LOTE motivational processes. In addition, the interplay of self- and identity-related orientations between English and the other TL was revealed to be potentially negative or positive when the main TL is more politically powerful than the other TL.

Research by De Smet et al. (2018), and De Smet et al. (2019) compared 896 French-native Belgian students studying either English or Dutch in a primary and secondary CLIL contexts (see [Section 3.5.3.1.2](#) for a discussion of this research specifically as it relates to motivation in CLIL). Students were divided into eight groups according to level (5th or 11th grade), learning context (CLIL or non-CLIL), and TL (English or Dutch). Collecting data by means of self-report questionnaires, De Smet et al. (2018) analysed learner anxiety and enjoyment, while De Smet et al. (2019) focused on language attitudes (perceived easiness and attractiveness of the TL) and motivation (expectancy for success, task value and cost). The former study indicated that while CLIL students reported significantly less anxiety than non-CLIL students overall, those in English CLIL reported significantly less anxiety and more enjoyment than students in Dutch CLIL. Similar results are reported in the latter study: more positive attitudes and higher motivation were reported by CLIL than non-CLIL students, and by those in English CLIL compared to Dutch CLIL. Crucially, effect sizes indicate that the TL, English or Dutch, plays a more crucial role in language attitudes and motivation than the educational approach does.

Geoghegan (2018) and Geoghegan and Pérez Vidal (2019) investigated the

motivation of undergraduate Spanish-Catalan bilinguals ($N = 68$), learning English as well as either German or French. Participants included both first- and second-year students, who took part or would take part in a compulsory study abroad (SA) period in their second year, choosing either to go to an English-speaking country, or a French-/German-speaking country. Geoghegan (2018) offered a quantitative study which compared these students cross-sectionally prior to and at the end of their SA period and compared students who sojourned in an English-speaking country with those in a German- or French-speaking country. Regarding the latter, results found a statistically significant difference between the two groups with regards to the Ideal L2 Self, with the English group reporting a higher overall mean. This suggests that the English group could better visualise themselves as the L2 user they wished to be than those in the French/German group. This study was followed up with a qualitative one by Geoghegan and Pérez-Vidal (2019) which focused on the effect of English as a lingua franca on the learners' motivation and identity. Following their SA, four participants from the sample above were interviewed using a semi-structured approach. Of these participants, two had sojourned in a French-speaking country (France and Belgium), one in a Germany, and one in England. Results highlighted a number of roles of English as a lingua franca, which were both positive and negative. Specifically regarding the effect of English on the L3, participants in non-English speaking countries reported that, at times, English could be used as a crutch, which helped them in times of L3 language-related difficulties. This was seen to have a positive effect on the learners' motivation to portray a multilingual self. On the other hand, English was also found to be an impediment in L3 acquisition. In some cases, participants reported consciously switching to their L2 to avoid difficulties in the L3. In other cases, native speakers of the L3 switched to English, much to the frustration of the students trying to practise their L3. As in the findings from Henry, discussed above, the authors highlight the importance of preparing students for these potentially detrimental effects on L3 motivation, so as to allow them to take full advantage of the learning situation.

Using the L2MSS framework, Calafato and Tang (2019) investigated the motivational self-concepts of 73 Arab teenagers who studied in English-medium schools and were learning an additional FL (French, Spanish, German, Japanese, Dutch, Italian, Russian, Turkish or Korean), with a focus in particular on gender differences. Participants completed a 54-item questionnaire in English, which included items on background and learning experience, motivation to be multilingual, ideal L3 self, ought-to L3 self, ideal

English self and ought-to English self. Results found that when comparing English and the students' L3, statistically significant correlations were found only between female participants' ought to-L3 and English selves.

Bui and Teng (2019) present a qualitative investigation with eight university students in Hong Kong (L1: Cantonese, L2: English and L3: Japanese). Using a dynamic systems theory perspective, two rounds of semi-structured interviews were carried out in order to analyse the learners' sources and changes of motivation and their self-perceptions of the L2 and L3 selves. Results highlighted the complexity of motivational dynamics in multilingual scenarios, in particular with regards to factors such as perceived ease and language status. For example, perceived ease of an L3 could reduce motivation for a more difficult L2, or the higher status of an L3 such as English could suppress the motivation for the L2.

Oakes and Howard (2019) used a quantitative approach to investigate the applicability of the L2MSS in LOTEs, specifically in French. Participants included 522 university learners of English ($n = 296$) and French ($n = 226$) in two European countries: Sweden and Poland. The questionnaires included seven motivational constructs, namely, the ideal L2 self, the ought-to L2 self, strong integrative orientation, weak integrative orientation, instrumental (promotion) orientation, intrinsic motivation and desire for proficiency. Results showed that there were statistically significant differences between French and English learners in four of the seven constructs: desire for proficiency and weak integrative, for which French learners had a higher mean, and instrumental and ought-to L2 self, for which English learners had a higher mean. The authors highlight in particular that the constructs under analysis play distinct roles in learners' motivational repertoires and are dependent on factors such as the TL, level of study, learning environment and length of study.

As can be seen, the complexity of L2 motivation has been found to be all the more intricate when an L3 is added to the mix. The above studies have demonstrated that these languages are by no means learned in isolation, but rather constantly interact with each other, a reality which has consequential implications for the learners' L2 and L3 motivation. In many cases, motivation towards English has been found to surpass L3 motivation, and even have a negative impact on it. However, some authors have also highlighted the possibility of counteracting this negative effect by adequately preparing students to successfully negotiate this interplay.

3.5. Research on Motivation and the Learning Context

This section provides an overview of L2 motivation in three different learning contexts: Formal Instruction (FI), Study Abroad (SA) and Content and Language Integrated Learning (CLIL). Studies dealing with FI and SA are discussed in order to provide a throughout background to motivation research in these important learning contexts. Those dealing with a CLIL context, however, are of particular importance to this thesis, given that it aims to investigate the potential benefits of this learning environment on learner motivation. Each of the following sections will first offer a brief theoretical introduction to L2 motivation in the specific learning context, and then outline some of the key research that has been carried out that area.

3.5.1. L2 Motivation in Formal Instruction

As is pointed out by Lamb (2017), despite increased mobility and access to FLs online, most individuals' first encounters with their TL occur in classrooms. This is inevitably an experience which may define attitudes and affect whether the student wishes to invest further in learning the language. In other words, a student who embarks on their language learning journey in the language learning classroom in a positive way may well be more likely to continue to make the effort to learn the language in the future.

Regarding research on L2 motivation in FI, Csizér (2017) pinpoints four key areas which provide empirical evidence: teachers, tasks, peers, and demotivation. Firstly, teacher motivation has been pinpointed as a vital factor, given that the successful use of motivation strategies is largely dependent on the teacher's motivation (Alderman, 2013). The employment of motivation strategies has then in turn been found to be related to student motivation as well as language achievement (Bernaus, Wilson & Gardner, 2009). Thus, a teacher who is motivated will be more likely to successfully incorporate strategies to motivate their students, which will in turn aid their students' language acquisition. Secondly, as discussed above, task motivation is another extremely important consideration in a classroom environment, given that it addresses why learners behave in a certain way in particular learning situations when they are carrying out a particular task (Dörnyei, 2002). Co-operative task environments, rather than individual or competitive ones, have been found to be particularly motivating for learners (Julkunen, 1989). In addition, L2 motivation can have a direct effect on how tasks are executed, with task motivation being co-constructed by the learners themselves (Dörnyei & Kormos, 2000; Dörnyei, 2002). Thirdly, the learner group is also understood to play a vital role in the

individual's L2 motivation (Busse & Walter, 2013). For example, Ghaith's (2003) study investigating how cooperative, individual and competitive learning modes affect the classroom climate indicated a significant relationship between cooperation and group cohesiveness. This finding is relevant in L2 motivation, given that the perceived cohesiveness of the group has also been found to affect learner motivation and to correlate with the learners' intended effort (Clément, Dörnyei & Noels, 1994). Finally, demotivation refers specifically to learners losing their L2 motivation over the course of the learning process. Demotivated students have been reported to have lower self-confidence, to not see the importance of learning the TL and to experience conflict with their teachers (Chambers, 1993). Sources of demotivation have most commonly been linked to teaching methods and learning tasks (Oxford, 1998; Ushioda, 1998).

Turning to recent research on L2 motivation, Table 3.5 offers a summary of some of the studies that have been carried out in these four areas over the last decade, which are then discussed in detail below.

Table 3.5

Research on L2 Motivation in FI

Area	Authors	Focus	Main Findings
Teacher Motivation	Ruesch et al. (2012)	Motivational strategies	Statistically significant differences between students and teachers' perceptions towards motivating strategies with regard to three macro strategies: Task, Effort and Comparison.
	Papi & Abdollahzadeh (2012)		Statistically significant relationship between teachers' motivational practices and students' motivated behaviour.
	Sugita McEown & Takeuchi (2014)		Students' proficiency and motivational intensity influence the perceived effectiveness of motivational strategies.
	Alrabai (2016)		Motivational intervention in the experimental group resulted in increased L2 motivation and higher achievement levels, as compared with students in the control group.

Table 3.5 (continued)

Task Motivation	Tankó & Csizér (2014)	Task performance on academic writing	There was a positive link between successful task completion and the participants' self-reported L2 motivation.
	Csizér & Tankó (2017)		
	MacIntyre & Serroul (2015)	Approach-avoidance motivation using a dynamic perspective	Motivation changed on a per-second basis throughout the tasks and was affected by the perception of task difficulty and the required vocabulary and grammar for successful task completion.
Group Dynamics	Chang (2010)	Group cohesiveness and group norms	Slight to moderate correlation were found between group processes and L2 motivation; the motivation of their classmates influences students' own motivation.
	Poupore (2016)	Group work dynamics, l2 motivation and language production	Statistically significant relationship between group work dynamic and task motivation (a stronger group work dynamic was associated with higher motivation) and between group work dynamic and language production (groups with stronger group work dynamics produced more language).
	Fukada et al. (2017)	Present Communities of Imagining	Past-self images, present-self images and future-self images become more positive over time when there are good group dynamics; participants' L2 motivation tends to resonate and harmonize the longer they are together.

Table 3.5 (continued)

Demotivation	Kim (2011)	English learning demotivation, private institutions	Statistically significant and consistent decrease in the participants' English-learning motivation; those who attended extracurricular English classes exhibited higher levels of instrumental and intrinsic motivation.
	Tanaka (2017)	Vocabulary acquisition in a demotivating environment	Perceived competence was a vital factor in motivating and demotivating learners; demotivated peers had a negative influence on learners' motivation; developing a larger vocabulary size was essential to enjoy and value learning.
	Kikuchi (2019)	Demotivation, L2MSS	Differing motivational dynamics were found in the four learners, with each one experiencing unique motivators and demotivators and reacting differently in each one; although the learners recognized their ideal and ought-to selves, these self-guides did not become stronger over time.

Firstly, teacher motivation and its effect on student outcomes has been one of the most highly investigated areas in L2 motivation in a classroom context. Many researchers, such as Ruesch, Bown and Dewey (2012), Papi and Abdollahzadeh (2012), Sugita McEown and Takeuchi (2014) and Alrabai (2016) have focused specifically on the use of motivational strategies. Ruesch et al.'s (2012) study investigated the perceptions of motivational strategies in 126 students and 30 teachers of Arabic, Chinese, French, Italian, Japanese, Russian and Spanish in a North American university context. The participants completed a questionnaire ranking 49 individual strategies on how motivating they found them. Results indicated significant differences between students' and teachers' perceptions towards motivating strategies with regard to three macro strategies: Task, Effort and Comparison. While teachers rated strategies related to Effort and Comparison significantly lower than students, students rated Task-related strategies significantly

higher than teachers, indicating important differences in student and teacher perceptions. Papi and Abdollahzadeh (2012) analysed the relationship between using motivational strategies and students' motivated behaviour in an Iranian context. Participants included 741 male secondary students and 17 EFL teachers, who were tested by means of the classroom observation instrument developed by Guilloteaux and Dörnyei (2008), a Student Motivational State questionnaire and Post-Lesson Teacher Evaluation scale which evaluated teachers' overall motivational practice. Results revealed a statistically significant relationship between teachers' motivational practices and students' motivated behaviour. However, as the authors point out, the data collected does not allow for a cause-and-effect interpretation. In other words, it cannot be stated whether it was the teachers' motivational practice which impacted their students' motivated behaviour, or vice versa. Sugita McEown and Takeuchi (2014) investigated the effect of the motivational strategies used by one instructor on the language learning of 222 undergraduate students at a Japanese university. Using a longitudinal approach, the study included four questionnaire sessions over the course of the semester, which dealt with the teacher's frequency of motivational strategies and students' motivation induced by each strategy. Student participants also completed an English proficiency test and motivational intensity towards general English learning questionnaire at the first data collection. This led to the grouping of students into two clusters: students with higher English proficiency and higher motivational intensity (Cluster A), and students with lower English proficiency and lower motivational intensity (Cluster B). Results revealed that while strategies such as starting the class on time and making a clear explanation for class assessments and exams showed positive correlations in both clusters, others such as speaking in a clear and loud voice or providing individual support for each student showed a positive correlation only in one cluster. These differences between these two groups indicate that students' proficiency and motivational intensity influence the perceived effectiveness of motivational strategies and should consequently be taken into consideration. Finally, Alrabai (2016) investigated the effects of motivational strategies on L2 motivation and English language achievement in Saudi Arabia. Participants included 437 learners, divided into experimental and control groups, and 14 EFL teachers. Four instruments were created for the study: one to identify the most valued motivational strategies in the context at hand, one to record teacher motivational practices and learner motivational behaviours in class, one to assess individual student motivational variables that could be affected or influenced by the teachers' motivational strategies, and one to

assess language and content achievement. The three instruments measuring motivational practices and levels were used both prior to and following the 10-week treatment period and learner achievement was measured three times during this period. The control group was exposed to six pre-selected motivational strategies through the course. Findings showed that the motivational intervention in the experimental group resulted in increased L2 motivation and higher achievement levels, as compared with students in the control group.

Secondly, research on task motivation in a classroom context includes that by Tankó and Csizér (2014), Csizér and Tankó (2017) and MacIntyre and Serroul (2015). The first two of these studies focused on the relationship between an academic writing task and L2 motivation in first-year English majors at a Hungarian university. Data was collected by means of a questionnaire addressing control strategies, dispositions toward L2 motivation, writing anxiety and self-efficacy beliefs, and an argumentative essay which was rated using the subscales task achievement, coherence and cohesion, grammar, and vocabulary. Results indicated a positive link between successful task completion and the participants' self-reported L2 motivation. MacIntyre and Serroul (2015) investigated approach-avoidance motivation during L2 task performance using a dynamic perspective. Participants included 12 undergraduate students attending Cape Breton University who completed a self-rating questionnaire on their French proficiency and a series of eight oral communicative tasks which were videotaped. Immediately after the tasks, participants' approach and avoidance motivation was also measured by having the participant watch their task performance and complete their idiodynamic ratings. This involved watching the task that they had just done and clicking the mouse to raise or lower the level of motivation shown on screen. Results indicated that motivation changed on a per-second basis throughout the tasks. Motivation was seen to be affected by the perception of task difficulty and the required vocabulary and grammar for successful task completion. A statistically significant positive correlation was also found between the participants' initial self-assessment of their task motivation and their actual motivation while performing the tasks.

Thirdly, the issue of group dynamics and motivation in FI has been addressed by Chang (2010), Poupore (2016) and Fukada, Murphey, Falout and Fukuda (2017). Chang (2010) explored the effect of variables such as group cohesiveness and group norms on students' motivation. To this effect, 152 Taiwanese university students completed a questionnaire addressing group related factors (group cohesiveness and group norms) and

L2 motivation (self-efficacy and autonomy). A subgroup of 12 students subsequently took part in semi-structured interviews. Quantitative results showed that there was a slight to moderate correlation between group processes and L2 motivation. These findings were further corroborated in the qualitative data, which revealed that students found that being in a class with motivated learners made them feel more motivated, whereas being with a group of stolid learners had the opposite effect. Poupore's (2016) study focused on the relationship between group work dynamics and L2 task motivation and language production. The participants were 10 Korean EFL learners enrolled in a Korean university TESOL program (intermediate to high intermediate proficiency). Throughout the course, the participants completed 15 different tasks in two work groups for each task. Data was collected by means of audio-video recordings, transcription of all verbal and non-verbal language produced in the tasks, a specifically designed group work dynamic measurement instrument, and post-task motivation questionnaires. Results revealed a statistically significant relationship between group work dynamic and task motivation, whereby a stronger group work dynamic was associated with higher motivation, and between group work dynamic and language production, whereby groups with stronger group work dynamics produced more language. Fukada et al. (2017) carried out a three-year panel study with 1351 students of English communication in different Japanese universities. Questionnaires used in the study investigated what the authors refer to as Present Communities of Imagining, which is an actively sharing and imagining classroom community. Within this community, each participant's three notional mind-time frames of L2 motivation are seen to interact among themselves and among those of others inside the classroom. Results indicated that given that there are good group dynamics, past-self images, present-self images and future-self images become more positive over time, with participants' L2 motivation tending to resonate and harmonize the longer they are together.

Finally, demotivation in a classroom context has been investigated by Kim (2011), Tanaka (2017) and Kikuchi (2017). Kim (2011) carried out a quantitative study with 6301 elementary school students from 3rd to 6th Grade in South Korea. Data was collected by means of a questionnaire based on Lamb (2007) investigating L2 motivation. School grades and prior experience in private English language institutions was also taken into consideration. Results showed a statistically significant and consistent decrease in the participants' satisfaction with their English-learning experience, expectations of ultimate success and intrinsic, extrinsic, integrative and instrumental motivation. In addition,

attendance in private institution was also found to be an important factor, with those who attended extracurricular English classes exhibiting higher levels of instrumental and intrinsic motivation. Tanaka (2017) focused on the role of motivation and peers on English vocabulary learning. Participants included 155 science and engineering students in a public technical college in Japan, who completed a VST and a L2 motivation questionnaire based on SDT and peer engagement/disengagement in learning. Key findings revealed that perceived competence was a vital factor in motivating and demotivating learners, and that while motivated peers had little impact on learners' motivation in a demotivating learning environment, demotivated peers had a negative influence. In addition, developing a larger vocabulary size was found to be essential to enjoy and value learning. Finally, Kikuchi's (2019) study involved four Japanese university students, majoring in international studies, who took part in a two-year study investigating changes in their L2 motivation. On a monthly basis, the participants were interviewed and also completed a 29-item questionnaire investigating the L2MSS. Results pinpointed differing motivational dynamics in the four learners, with each one experiencing unique motivators and demotivators and reacting differently in each one. In addition, although the learners recognized their ideal and ought-to selves, these self-guides did not become stronger over time.

To conclude, motivation in a FI context has largely been investigated with regard to teacher motivation, task motivation, group dynamics and demotivation. While research has shown the benefits of teachers' use of motivational strategies, differences have also been revealed between students' and teachers' perceptions of them (Ruesch et al., 2012) and between students with different levels of proficiency and motivational intensity (Sugita et al., 2014). Regarding task motivation, the research discussed above highlights the link between successful task completion and L2 motivation (Tankó & Csizér, 2014; Csizér & Tankó, 2017), but also highlights the need to approach this research using a dynamic perspective due to the per-second changes that have been found during task completion (MacIntyre & Serroul, 2015). Group dynamics in the classroom have also been found to play a key role in L2 motivation, with research indicating that a positive group dynamic has a beneficial effect on L2 motivation (Poupore, 2015). In addition, peers' motivation has also been seen to play a key role, with students feeling more motivated themselves when in a group with other motivated students (Chang, 2010). Finally, demotivation research has shown that learners often lose motivation over the course of year, though this may be counteracted to some extent by having additional,

extramural contact with English, such as private English classes (Kim, 2011). Kikuchi's (2019) qualitative research again points to varying motivational dynamics, highlighting the differences in each individual's learning experience. Moving forward, Csizér (2019) highlights the need for both quantitative and qualitative research, as well the need to investigate the relationship between classroom-related variables and L2 motivation.

3.5.2. L2 Motivation in Study Abroad

According to Yim, Clément and MacIntyre (2019, p. 235), a natural learning context is often considered to be superior to a classroom one, given that it “promotes motivation by presenting the learner with more tangible language achievements and bolstering their feeling of autonomy”. In addition, intercultural contact has also been found to have an impact of L2 motivation (Kormos, Csizér & Iwaniec, 2014; Dörnyei & Csizér, 2005), given that it creates opportunities to develop language skills and assists in shaping the learner's attitudinal and motivational disposition. A rich, authentic learning environment such as SA may not only be beneficial for the learner's motivation, but has also been suggested to potentially lower language learning anxiety (Gabiella Morreale, 2011). However, while the possibility of going abroad will evidently be a source of motivation for many students, for those who do not have the opportunity, this may be a direct source of demotivation. As Ryan and Mercer (2011) point out, students who study only in a classroom environment in their home country may feel that they lack legitimacy because they have not and will not have the chance to be tested or authenticated in the TL country. In addition, learners who do go abroad may not always be able to take advantage of the learning opportunities, as in some cases motivational and attitudinal deficits can prevent learners from interacting with TL speakers in the first place (Isabelli García, 2006).

With regards to research on L2 motivation in a SA context, some studies have focused solely on the SA context, analysing learner motivation with regard to a number of different issues such as social networks, perceived learning progress, or language ability. Other studies have quite often been comparative in nature, contrasting learners in a traditional classroom context with those in SA, so as to verify the purported benefits of one context over the other. A selection of these studies is summarised in Table 3.6 and is discussed in turn below.

Table 3.6*Research on L2 Motivation in Study Abroad*

	Authors	Focus	Main Findings
Study Abroad	Isabelli-Garcia (2006)	Motivation, attitudes and SA social networks	Motivation was influenced by how successfully students incorporated themselves into social networks.
	Badstübner & Ecke (2009)	Student Expectations, Motivations, TL Use, and Perceived Learning Progress	Participants' expectations were significantly higher than their perceived progress in all proficiency areas except cultural learning.
	Allen (2010)	Language-learning motivation and orientations	Two key orientations motivated participants to learn French at college level: linguistic motives and career-oriented motives.
	Sasaki (2011)	Length of stay, L2 Writing Ability and Motivation	Only students who had spent more than 8 months abroad became more intrinsically motivated and voluntarily practised to improve their L2 writing.
	Irie & Ryan (2014)	L2 learner concept	Learners may go through difference states during their SA experience, such as <i>naïve optimist</i> , <i>shell-shocked doubters</i> or <i>comfortable users</i> .
	Martinsen et al. (2014)	Perceived foreign accent, level of instruction, motivation	Level of instruction rating explained around 67% of the variance in pronunciation rating scores, and around 68% of the variance when motivation was included as a predictor.

Table 3.6 (continued)

Study Abroad vs. Formal Instruction	Hernández (2010)	Motivation, interaction and speaking proficiency	Groups had similar motivation profiles; both motivation and interaction were found to be important factors for developing speaking skills in both learning contexts.
	Mora & Valls-Ferrer (2012)	Oral Fluency, Accuracy, and Complexity	Clear gains in oral production found after SA but not after FI; learners whose fluency improved the most during the SA period consistently reported very high motivation.
	Trenchs-Parera & Juan-Garau (2014)	Motivation and beliefs	FI led to positive changes in self-confidence, intended effort and willingness to speak in public; SA strengthened the development of positive motivational substrates.
	Geoghegan (2018)	Motivation and identity in L2 or L3 study abroad	Statistically significant differences international posture, WTC and interest in FLs.

Of the above studies, Isabelli-Garcia (2006), Badstübner and Ecke (2009), Allen (2010), Sasaki (2011), Irie and Ryan (2014) and Martinsen, Alvord and Tanner (2014) investigate SA without comparing it to other learning contexts. Isabelli-Garcia (2006) analysed the motivation, attitudes and SA social networks of four students of Spanish from the United States using a mixed methods study. Data collection included pre- and post- program oral proficiency interviews, informal interviews, diary entries for motivation/attitude orientation and social network contact logs. Results showed that learners’ continued motivation was influenced by how successful they were in incorporating themselves into social networks, thus stressing that SA experiences vary from learner to learner and highlighting the importance of taking other factors into consideration. Badstübner and Ecke (2009) focused on student expectations, motivations,

use of the TL and perceived learning progress in a German context. Participants included 23 US students enrolled in the one-month Summer Study-in-Leipzig-Germany Program; five of whom were in the second-year course and eighteen of whom were in the third-year course of the program. Data was collected by means of two sets of questionnaires which were administered at the start and end of the program. While the first addressed students' goals and motives for participating in the program and language learning expectations, the second had students rate their gains across seven proficiency areas. Results showed that the participants' expectations were significantly higher than their perceived progress in all proficiency areas except cultural learning. In addition, there were statistically significant correlations between the frequencies of L2 listening and gains in L2 listening proficiency, and between the goals "to be in contact with Germans" and "cultural enrichment" with participants' expected gains in speaking and listening skills. Allen (2010) carried out a qualitative study with six American intermediate-level students of French who took part in a six-week SA. Multiple instruments were used to collect data on language-learning motivation, including questionnaires, SA application essays, e-mail interviews, and learning blogs. The results demonstrate two key orientations which motivated the participants to learn French at college level, namely linguistic motives and career-oriented motives. Partaking in a SA was also viewed as a critical step in achieving linguistic objectives. Sasaki (2011) investigated the differences in the length of SA on L2 writing and motivation. Participants included 37 Japanese students studying abroad for a period of 1.5 to 11 months in English-speaking countries, who were tested both at the beginning of their first year and also in the middle of their second, third, and fourth years of university. The students were divided into four groups; one group of students who did not go abroad and three groups of students who differed in the number of months spent abroad: 1.5-2, 4 or 8-11 months. At each testing period, participants completed an argumentative composition and, following this, an interview regarding their L2 writing strategies, L2 classes and motivation. Participants also took part in an additional interview four months after the final compositions in which they reflected on the changes in their L2 writing and motivation over the three-and-a-half-year period. Regarding motivation, results indicated that only students in the fourth group, who had spent more than 8 months abroad, became more intrinsically motivated and voluntarily practised to improve their L2 writing. Irie and Ryan (2014) investigated the dynamic nature of L2 motivation in SA with regards to learner L2 self-concept using a longitudinal approach. A total of 19 participants took part in both rounds of the data collection, during which they had to sort

50 cards with statements concerning beliefs about English into three piles: statements which they generally agreed with, did not agree with, or felt relatively neutral about. They then had to rank the statements from -5 to 5 with regards to how descriptive the statement was about their views. The results include the recognition of different states which learners may experience at the start of the SA. For example, before going abroad, learners often fell into the category of *naïve optimist*, having a keen sense of anticipation and excitement about embarking. However, after arriving in the host country, learners often became *shell-shocked doubters*, whereby their sense of anticipation was removed, and the reality of the situation shifted them into a volatile state. Other states such as the *comfortable user* were also marked as dependent on support from the SA environment, friends made and feelings of achievement. Lastly, Martinsen et al. (2014) analysed the connection between motivation, cultural sensitivity, and level of instruction with perceived foreign accent. Participants ($N = 102$) came from six different levels of Spanish classes at university. This included 12 beginners, 26 intermediates, 16 high intermediates, two groups of third year students (10 who had no experience abroad and 26 who had completed two months of intensive language training followed by a 22-month period abroad) and 9 fourth year students. All participants completed a survey on motivational intensity based on the AMTB and took part in an oral proficiency interview. Results indicated that the level of instruction rating explained around 67% of the variance in pronunciation rating scores. When motivation was included as a predictor, the model explained approximately 68% of the variance.

In addition to these studies, Hernández (2010), Mora and Valls-Ferrer (2012), Trenchs-Parera and Juan-Garau (2014) and Geoghegan (2018) offer a comparison of students in SA and in FI. Hernández (2010) compared at home and SA students in terms of how their motivation and interaction shaped their speaking proficiency. Forty-four native English speakers took part in the study. They were all Spanish language learners who had taken a minimum of four semesters of college Spanish, or equivalent, and were divided into two groups: an SA group ($n = 20$) and an at home group ($n = 24$). Both groups completed a motivation questionnaire, a language contact profile and pre-test and post-test simulated oral proficiency interview. Results revealed that both groups had similar motivation profiles, though the SA group reported higher use of the TL outside of class and also improved their speaking proficiency more than the at home group. Both motivation and interaction were found to be important factors for developing speaking skills in both learning contexts. Mora and Valls-Ferrer (2012) also focused on oral

production skills, analysing the fluency, accuracy, and complexity of 30 Catalan-Spanish undergraduate learners of English both at home and in an SA period, and a control group of 10 native-English speakers. The study used a longitudinal approach, assessing students three times: upon enrolment at the university, after two 3-month terms of FI, and immediately after a 3-month SA period in an English-speaking country. At each data collection, participants took part in a pair interview to assess their oral production skills. They also completed three questionnaires addressing their language background and use, motivation, beliefs, and attitudes in learning English and the TL culture, and the SA. While clear gains in oral production were found after SA, no such improvement was found after FI. In addition, learners whose fluency improved the most during the SA period consistently reported very high motivation, indicating the possible connection between motivational variables and improvement in speaking proficiency during SA. Trenchs-Parera and Juan-Garau's (2014) study had a similar profile to that above, but focused more specifically on motivation and learner beliefs. The participants were 70 Catalan-Spanish undergraduate learners of English who were tested four times: upon enrolment at the university, after two 3-month terms of FI, immediately after a 3-month SA period in an English-speaking country, and after a period of 15 months following the SA. At each stage, participants answered a 48-item self-report questionnaire investigating L2 motivation and language learner beliefs. Results indicated that FI led to positive changes in self-confidence, intended effort and willingness to speak in public. On the other hand, SA was seen to strengthen the development of positive motivational substrates, such as instrumentality, integrativeness, attitudes towards learning English, a reduction in anxiety and an increase in the attribution of importance of listening abilities. Finally, as discussed above, Geoghegan (2018) compared Spanish-Catalan bilinguals in either the first or second year of their undergraduate degree in a cross-sectional study, prior to and at the end of their three-month SA period. Results indicated differences in the two groups only with regard to some categories, such as international posture, WTC and interest in FLs. Surprisingly, the first-year students, prior to SA, had a statistically significant higher interest in FLs than the second-year students at the end of SA. In addition, the first-year students had a higher WTC in their native language than the second-year students, whereas there was no statistically significant difference between the groups in terms of WTC in the TL. On the other hand, regarding international posture, second-year students were significantly more likely to have thoughts they wished to share with others of different nationalities than the first-year students.

In summary, many studies have attested the benefits that a SA period has on L2 motivation. However, it should also be cautioned the effectiveness of SA will depend very much in the exact nature of the experience. For example, Pérez-Vidal (2014) highlights eight key features of SA: length of SA programme, accommodation and living conditions, having a job, set pre-departure language level, pre-departure preparation, point in the curriculum, academic work assignments abroad, and re-entry conditions. These are factors which will evidently affect the SA experience, and which should consequently be taken into consideration when evaluating the efficiency of SA as a learning context.

3.5.3. L2 Motivation in CLIL

Despite being one of the most widely researched topics in SLA, motivation has received considerably less attention within a CLIL context. Lasagabaster (2019) attributes this sidelining of motivation to a focus on other factors, such as FL competence and the development of students' L1 and content learning. Notwithstanding, high levels of student motivation are repeatedly put forward as a reason for adopting a CLIL approach (Pavesi, Bertocchi, Hofmannová & Kazianka, 2001; Ruiz de Zarobe, 2011; Attard-Montalto & Walter, 2021). As Doiz et al. (2014a) point out, given that motivation has been found to be one of the most influential benefits in SLA, the advantages in a CLIL approach are simply taken for granted, despite a dearth of empirical studies confirming the benefit.

Lasagabaster (2019) highlights three main reasons why CLIL, in theory, may have a positive influence on motivation:

1. It provides a cognitively challenging situation.
2. It may promote fruitful discussion.
3. It provides a sense of ownership in teaching and learning.

Firstly, according to Coyle's (1999) four Cs curriculum (see also [Section 5.1.2](#)), CLIL has a number of general parameters, namely Content, Communication, Cognition, and Culture. The framework centres on "the interrelationship between content (subject matter), communication (language), cognition (learning and thinking) and culture (social awareness of self and 'otherness')" (Coyle, 2007, p. 550). Thus, one of the main principles in the approach is that CLIL should cognitively challenge learners (Wiesemes, 2009). This cognitively challenging environment has in turn been linked to meaningful use of

the TL and an improved sense of achievement (Lasagabaster, 2019). In addition, CLIL may encourage rich discussion on pedagogical issues and practices, in turn leading to greater motivation. Finally, CLIL may give learners a sense of control, by allowing them the freedom to use the language according to their needs in their content-focused tasks. For example, in Cross (2013), using language as a means to learn content provided students with greater ownership and creativity in the language they used and allowed them to set their own rules for how to do so.

Following the initial dearth of motivation research in CLIL, there has been a burst of studies particularly in the last five years which aim to confirm these purported benefits of a CLIL approach on motivation. According to Lasagabaster (2019) these studies can largely be grouped into three types: those which quantitatively or qualitatively measure students' actual motivation, those which focus on multilingualism in CLIL contexts, and those which analyse stakeholders' beliefs. The following two subsections provide an overview of some of the key research that has been carried out, focusing first on the quantitative and qualitative measurement of CLIL in monolingual and multilingual contexts, and then dealing with studies which have addressed stakeholders' beliefs.

3.5.3.1. Measurement of Motivation in CLIL. Quantitative and qualitative research has by far been given the most attention in CLIL motivation research. Given the extent of this research, it will be discussed below in two further subsections, focusing first on contributions in a Spanish context and then outlining other important research elsewhere in Europe.

3.5.3.1.1. Measurement of Motivation in CLIL in Spain. Spain has been a forerunner in CLIL research, a fact which is particularly noticeable in studies addressing motivation in CLIL. Some of this key research is summarised in Table 3.7 and discussed below.

Table 3.7

Quantitative and Qualitative Measurement of Motivation in CLIL in Spain

Authors	Age group	Focus	Main Findings
Lasagabaster (2011)	Teenagers (Mean age 15) 164 CLIL + 27 non-CLIL	Relationship between motivation and language proficiency	Statistically significant differences between the two groups for both motivation and language proficiency to the advantage of the CLIL group.
Fernández Fontecha & Canga Alonso (2014)	Young learners (4 th grade) 31 CLIL + 31 non-CLIL	Motivation and gender	All participants more extrinsically than intrinsically motivated; the non-CLIL group had higher motivation than the CLIL group both in terms of intrinsic and extrinsic motivation, and overall motivation.
Doiz et al. (2014a)	Teenagers (7 th grade) 107 CLIL + 101 non-CLIL Teenagers (9 th grade) 114 CLIL + 71 non-CLIL	Motivation and the variables of age, sex and parental socio-cultural level	CLIL students were more motivated, particularly with regards to intrinsic motivation, instrumental orientation and interest in FLs/cultures.
Doiz et al. (2014b)	Teenagers (7 th grade) 107 CLIL Teenagers (9 th grade) 114 CLIL	Qualitative assessment of what (de)motivates students	Several issues need to be taken into consideration, given the variability of students in CLIL classes.

Table 3.7 (continued)

Fernández Fontecha (2015)	Young learners (4 th grade) 58 CLIL Teenagers (7 th grade) 304 non-CLIL	Motivation, gender and vocabulary breadth	The majority of students in both groups were highly motivated, the primary group slightly more so, though this result was not statistically significant; no relationship was found between motivation and receptive vocabulary knowledge in the 4 th grade students; a small positive significant correlation was found between the two for the 7 th grade students.
Lasagabaster & López Beloqui (2015)	Young learners (5 th grade) 23 Book-based CLIL 32 Project Work CLIL 32 non-CLIL	Motivation and type of approach (Book-based versus Project Work)	Statistically significant differences were found for intrinsic and integrative motivation the book-based CLIL and non-CLIL groups, to the advantage of the CLIL group, and between the book-based and project work CLIL groups, to the advantage of the book-based group.
Heras & Lasagabaster (2015)	Teenagers (10 th grade) 25 CLIL + 21 non-CLIL	Affective factors, vocabulary and gender	The CLIL group had higher scores in all four shared clusters (Instrumental, Ideal L2 Self, Ought-to L2 Self and Self-esteem in EFL class), though the results were not statistically significant.

Table 3.7 (continued)

Arribas (2016)	Teenagers (10 th grade) 73 CLIL + 19 non-CLIL	Motivation and receptive vocabulary	Positive correlations between attitudes and results in each VLT frequency band were found for all students, regardless of instruction type.
Lasagabaster & Doiz (2017)	Teenagers (7 th grade) 97 CLIL + 61 non-CLIL Teenagers (9 th grade) 102 CLIL + 44 non-CLIL	Longitudinal impact of CLIL on affective factors	CLIL and non-CLIL groups' means were very similar; no downward motivational trend found in non-CLIL students; CLIL did not help to sustain the students' motivation over the two- or three-year period; motivation to learn the content was maintained in the CLIL classes.
Navarro Pablo (2018)	Young learners (6 th grade) 61 CLIL + 133 non-CLIL Teenagers (10 th grade) 75 CLIL + 83 non-CLIL	Affective factors Interaction of motivation and language attainment	CLIL students were more motivated than non-CLIL students; motivational variables had a statistically significant effect on the differences in the subtests.
Pladevall- Ballester (2019)	Young learners (5 th to 6 th grade) 138 CLIL + 149 non-CLIL	Development of language learning motivation over two years	CLIL learners' motivation towards the L2 Learning Experience increased over time, though this is not the case for non-CLIL learners.

Firstly, Lasagabaster (2011) analysed the relationship between motivation and language proficiency in 164 CLIL and 27 non-CLIL teenagers with an average age of 15. Instruments included a 13-item questionnaire to assess motivation, the Oxford Placement

Test to assess grammar and listening skills, a letter to an English family to assess writing, and the frog story (Mayer, 1969) to assess speaking. Findings suggested that CLIL is beneficial for both motivation and language proficiency, with statistically significant differences between the two groups.

Fernández Fontecha and Canga Alonso (2014) compared the motivation of 31 CLIL and 31 non-CLIL 4th grade learners using an adapted version of Gardner's (1985) AMTB. Results indicated that all subjects were more extrinsically than intrinsically motivated, but that the non-CLIL group had higher motivation than the CLIL group both in terms of intrinsic and extrinsic motivation, as well as in overall motivation. Results also pointed to greater variation in the motivation scores of CLIL learners.

Doiz et al. (2014a, 2014b) presented their findings on the role of motivation and attitudes towards learning English in students in the Basque Country in Spain. Participants in the studies included 107 CLIL and 101 non-CLIL 7th grade students and 114 CLIL and 71 non-CLIL 9th grade students who were assessed using a previously piloted and validated questionnaire with both closed- and open-ended questions. While Doiz et al. (2014a) investigated the relationship between motivation and the variables of age, sex and parental socio-cultural level of these four groups, Doiz et al. (2014b) carried out a qualitative assessment of the motivation of the two CLIL groups. The former indicated that the CLIL students were more motivated, particularly with regards to intrinsic motivation, instrumental orientation and interest in FLs/cultures. The latter offered more detailed insight into what students find motivating and demotivating about the CLIL approach. The findings offer several practical issues to consider and point to the variety of students in the classes, highlighting the need to cater to these differences.

Fernández Fontecha (2015), discussed in the previous section with regards to vocabulary, also analysed the motivation of the 4th grade CLIL and 7th grade non-CLIL learners by means of a questionnaire adapted from Gardner's (1985) AMTB. The majority of students in both groups were found to be highly motivated, with the primary group slightly overcoming the secondary students, though this result was not statistically significant. In addition, though no relationship was found between motivation and receptive vocabulary knowledge in the 4th grade students, a small positive significant correlation was found between the two for the 7th grade students.

Lasagabaster and López Beloqui (2015) compared 55 CLIL students, 23 of whom followed a book-based methodology and 32 of whom followed a project work methodology, and 32 non-CLIL students. Students' motivation was assessed by means of

a questionnaire containing a number of items concerning intrinsic, extrinsic, instrumental and integrative motivation, as well as interest in other cultures and the learning environment. Upon comparing the book-based CLIL and non-CLIL groups, statistically significant differences were found for intrinsic and integrative motivation, with the CLIL group having a higher mean. As for the book-based and project work CLIL groups, statistically significant differences were again found for intrinsic and integrative motivation, with the book-based group reporting higher motivation in each case.

Heras and Lasagabaster (2015) assessed 25 CLIL and 21 non-CLIL students using a questionnaire on motivation and self-esteem which investigated five clusters: Instrumental, Ideal L2 Self, Ought-to L2 Self, Self-esteem in EFL class and Self-esteem in PE class (this final cluster was relevant only to the CLIL students who were taking PE through English). Results showed that although the CLIL group had higher scores in all four shared clusters (Instrumental, Ideal L2 Self, Ought-to L2 Self and Self-esteem in EFL class), the results were not statistically significant. The authors highlight the importance of the intensity of the CLIL programme, given that the low-intensity programme in which the participants in this study were enrolled did not lead to any significant differences in the students' motivational stance. The value of intensity has also been corroborated more recently in a longitudinal study by Merino and Lasagabaster (2018b), who noted that while no difference in English proficiency was found between non-CLIL and lower exposure CLIL groups, statistically significant differences were observed in CLIL groups with higher exposure to the TL.

In addition to assessing vocabulary, Arribas (2016) also investigated the 73 CLIL and 19 non-CLIL students in terms of their motivation towards English, analysing this alongside their receptive vocabulary. For all students, regardless of instruction type, there was a positive correlation between attitudes and results in each VLT frequency band, indicating that the higher the students' motivation, the higher their level of receptive vocabulary. However, while CLIL students had a higher score in their attitude towards English, the results were not statistically significant.

Lasagabaster and Doiz (2017) carried out a longitudinal study investigating the impact of CLIL on affective factors. Participants included 97 CLIL and 61 non-CLIL students in 7th grade and 102 CLIL and 44 non-CLIL in 9th grade, assessed using a questionnaire based on scales already used by Gardner (1985) and Schmidt and Watanabe (2001) and carried out over a three-year (younger students) or two-year (older students) period. Results showed that the means of the CLIL and non-CLIL groups were very

similar, perhaps, as the authors suggest, due to the hegemonic position of English. In addition, and contrary to expectations, no downward motivational trend was found in non-CLIL students, and CLIL did not help to sustain the students' motivation over the two- or three-year period. However, motivation to learn the content was maintained in CLIL.

Navarro Pablo (2018) compared the interaction of motivation and language attainment of 61 CLIL and 133 non-CLIL students in 6th grade and 75 CLIL and 83 non-CLIL students in 10th grade. Participants came from eight different centres, three of which were rural and five of which were urban. Potential participants were tested in advance in order to match CLIL and non-CLIL students in verbal intelligence and motivation, excluding students with higher and lower scores from the original sample. Instruments included a specifically designed language proficiency test (measuring Use of English, Vocabulary, Listening, Reading, Speaking Total, Grammar, Lexical Range, Fluency Interaction, Pronunciation and Task Fulfilment) and Pelechano's (1994) MA test to measure motivation. Results indicated differences in the educational levels analysed. While at primary level the CLIL group outperformed the non-CLIL group in all subtests, this was not statistically significant for the subtests Listening and Reading. At secondary level, however, CLIL students outperformed non-CLIL students on all subtests, differences which were statistically significant in all cases. Results also showed that CLIL students were more motivated than non-CLIL students, and that motivational variables had a statistically significant effect on the differences in the subtests.

Pladevall-Ballester (2019) followed 138 CLIL and 149 non-CLIL students from 5th to 6th grade, using a longitudinal approach to track the development of language learning motivation. Results of the motivation questionnaire used reveal that, unlike their non-CLIL peers, CLIL learners' motivation towards the L2 Learning Experience increased over time.

3.5.3.1.2. Measurement of Motivation in CLIL in Europe. In addition to the above studies carried out in a Spanish context, other notable contributions in Europe come from Merisui-Storm (2007) and Seikkula-Leino (2007) in Finland, Verspoor, de Bot and Xu (2015) in the Netherlands, Sylvén and Thompson (2015) in Sweden, Rumlich (2016, 2017) in Germany, Pfenninger (2016) in Switzerland, Otwinowska and Foris (2017) in Poland, Young (2018) in Scotland and De Smet et al. (2018, 2019) in Belgium. This selection of research is summarised in Table 3.8 and discussed below.

Table 3.8*Quantitative and Qualitative Measurement of Motivation in CLIL in Europe*

Authors	Age group	Focus	Main Findings
Merisuo-Storm (2007)	Young learners (4 th grade) 70 CLIL + 75 non-CLIL	Attitudes towards FL learning, reading and writing Gender	CLIL students had statistically significantly more positive attitudes towards their language learning than their non-CLIL peers.
Seikkula-Leino (2007)	Young learners (5 th + 6 th grade) 116 CLIL + 101 non-CLIL	Achievement and affective factors	CLIL students had a low self-concept in the FL in relation to their non-CLIL peers but had a strong motivation to learn.
Verspoor et al. (2015)	Teenagers (7 th grade) 83 CLIL, 64 non-CLIL + 49 control Teenagers (9 th grade) 74 CLIL, 68 non-CLIL + 41 control	Development of English proficiency Out of school contact Motivation/Attitude	CLIL students outperformed non-CLIL students; scholastic aptitude and initial proficiency were strong predictors 7 th grade, though in 9 th grade scholastic aptitude no longer played a role but initial proficiency and motivation/attitude did.
Sylvén & Thompson (2015)	Teenagers (10 th grade) 109 CLIL + 68 non-CLIL	L1, motivation and gender	CLIL outperformed non-CLIL in interest in FLs, motivation, ideal L2 self, self-confidence and WTC in English; non-CLIL students were also more ethnocentric and had higher English anxiety.

Pfenninger (2016)	Teenagers (12 th grade) 100 CLIL + 100 non-CLIL	L3 proficiency, starting age and motivation	Although starting age did not affect motivation, the type of instruction had a statistically significant impact, enhancing motivation.
Rumlich (2016, 2017)	Teenagers (7 th to 9 th grade) 953 pupils	English proficiency Self-concepts and interest outcomes	Large differences prior to CLIL as a result of selection, preparation, and class composition effects; no CLIL-related benefits for general English proficiency or interest in English classes, and only a small increase in CLIL-attributable EFL self-concept.
Otwinowska & Foris (2017)	Young learners (4 th grade) 65 CLIL Young learners (5 th grade) 75 CLIL	Affectivity and cognition	Negative affectivity and grades in science and mathematics (but not English) were significant predictors of Intellectual Helplessness in CLIL.
Young (2018)	Young learners (5 th + 6 th grade) 18 CLIL	Motivation towards FL learning	The CLIL approach provided an engaging, motivating and enjoyable language learning environment for students.

Table 3.8 (continued)

<p>De Smet et al. (2018)</p>	<p>Young learners (5th grade) 276 CLIL + 165 non-CLIL Teenagers (11th grade) 240 CLIL + 215 non-CLIL</p>	<p>Comparison of anxiety and enjoyment in the classroom in two TLs (English and Dutch)</p>	<p>CLIL students reported significantly less anxiety than non-CLIL students overall, those in English CLIL reported significantly less anxiety and more enjoyment than students in Dutch CLIL.</p>
<p>De Smet et al. (2019)</p>		<p>Language attitudes and motivation in two TLs (English and Dutch)</p>	<p>More positive attitudes and higher motivation were reported by CLIL than non-CLIL students, and by those in English CLIL compared to Dutch CLIL; TL plays a more crucial role in language attitudes and motivation than CLIL.</p>

Firstly, Merisuo-Storm (2007) analysed attitudes toward FL learning, reading and writing in 70 CLIL and 75 non-CLIL 4th grade students in Finland. The instrument used was based on McKenna and Kear’s (1999) Elementary Reading Attitude Survey and Kear Coffman, McKenna and Ambrosio’s (2000) Writing Attitude Survey. Results indicated that CLIL students had statistically significantly more positive attitudes towards their language learning than their non-CLIL peers.

Also in a Finnish context, Seikkula-Leino (2007) compared 116 CLIL and 101 non-CLIL students in 5th and 6th grade in terms of achievement and affective factors. Instruments included Raven’s (1989) non-verbal intelligence test, Wechsler’s (1949) vocabulary test to measure verbal intelligence, tests concerning mathematics and Finnish as a mother tongue and likert-scale tests of self-esteem, learning self-concept and FL self-concept. Regarding affective factors, results found that although CLIL students had a low self-concept in the FL in relation to their non-CLIL peers, they had a strong motivation to learn.

In the Netherlands, Verspoor et al. (2015) carried out a longitudinal study investigating the development of English proficiency, language contact, and motivation/attitudes. Students were divided into three groups for each grade based on three educational streams: the bilingual stream (CLIL), the regular stream (regular non-CLIL) and the gymnasium stream (control non-CLIL) (see Verspoor et al., 2015, for an overview of the Dutch secondary system). Participants thus included three groups of 7th grade students (83 CLIL, 64 regular non-CLIL and 49 control non-CLIL) and three groups of 9th grade students (74 CLIL, 68 regular non-CLIL and 41 control non-CLIL). The students were tested three times in one academic year and assessed based on their CITO scores (an independent assessment of Dutch school pupils) and a language contact, motivation and attitudes questionnaire developed by Berns, de Bot and Hasenbrink (2006). Results indicated that CLIL students outperformed students in the non-CLIL groups, and pinpointed a dynamic interplay between condition and factors such as initial proficiency, scholastic aptitude, out of school contact, and motivation/attitude, as proficiency increased. While in 7th grade scholastic aptitude and initial proficiency were found to be strong predictors, in 9th grade scholastic aptitude no longer played a role but initial proficiency and motivation/attitude did.

In Sweden, Sylvén and Thompson (2015) investigated the motivation of 109 CLIL and 68 non-CLIL 10th grade students across three different schools using Ryan's (2009) MFQ on the L2MSS. CLIL students were found to be more interested in FLs and more positive towards learning English, and have a stronger *Ideal L2 Self*, more English self-confidence and a higher WTC in English. Non-CLIL learners were more ethnocentric and had higher English anxiety.

In Switzerland, Pfenninger (2016) analysed the association between L3 performance and starting age, motivation, and type of instruction, comparing 100 CLIL and 100 non-CLIL 12th grade students by means of a 15-item likert-type questionnaire (Pfenninger & Singleton, 2017). Participants came from 12 classes in five schools and were divided into four groups of 50 students according to starting age and learning constellation in primary and secondary school: Early CLIL, Early Non-CLIL, Late CLIL and Late Non-CLIL. Given their different learning experiences, the number of hours of exposure varied across groups from 1770 hours for the Early CLIL students, 1330 hours for the Late CLIL students, 1170 hours for the Early Non-CLIL students and 730 hours for the Late Non-CLIL students. Results indicated that although starting age did not affect motivation, the type of instruction has a statistically significant impact, enhancing

motivation by around 0.40 ± 0.10 points on a 5-point scale.

In Germany, Rumlich (2016, 2017) conducted a longitudinal study which investigated the effects of CLIL on general EFL proficiency, EFL self-concept and interest over a period of two school years (German Year 6 to Year 8). Participants included 993 pupils, 503 CLIL and 473 non-CLIL, in 43 different classes., who were assessed on general EFL proficiency by means of a collection of validated C-test and EFL self-concept and interest by means of a number of a four-point Likert scale items. Results highlight large differences prior to CLIL, as a result of selection, preparation, and class composition effects. Following the two-year period, no CLIL-related benefits were found for general English proficiency or interest in English classes, and only a small increase in EFL self-concept which could be attributable to CLIL.

In Poland, Otwinowska and Foris (2017) investigated the relationship between affectivity and cognition in 4th grade ($n = 65$) and 5th grade ($n = 75$) CLIL students, using an attitude survey and the Intellectual Helplessness scale (Sędek, 1995). Results found negative affectivity and grades in science and mathematics (but not English) to be significant predictors of Intellectual Helplessness in CLIL.

Young (2018) presents a small study which addresses the language learning motivation of 18 5th and 6th grade Scottish students learning Italian as an L2, analysed by means of focus groups, questionnaires and a reflective journal. Findings indicate that the CLIL approach provided an engaging, motivating and enjoyable language learning environment for students.

Finally, and of central importance to the current work given its focus on multilingual CLIL, De Smet et al. (2018, 2019), discussed above in Section 3.4.2.2 with regard to research comparing motivation towards English and LOTEs, carried out an analysis of 896 CLIL and non-CLIL pupils in primary and secondary education in Belgium. Of particular interest in these studies is the fact that they analyse the motivational differences of French speaking students learning two difference FLs: English and Dutch. Participants thus included eights groups: English CLIL ($n = 102$) and non-CLIL ($n = 97$) and Dutch CLIL ($n = 174$) and non-CLIL ($n = 68$) in 5th grade, and English CLIL ($n = 100$) and non-CLIL ($n = 102$) and Dutch CLIL ($n = 140$) and non-CLIL ($n = 113$) in 11th grade. De Smet et al. (2018) measure, among other variables, pupils' anxiety and enjoyment in the classroom by means of an extensive self-report questionnaire. CLIL students overall reported significantly less anxiety than non-CLIL students. In addition, those in English CLIL reported significantly less anxiety and more

enjoyment than students in Dutch CLIL. Regarding educational level, results also indicate that while the 5th grade students reported stronger emotions, the effects of CLIL and English were much larger in the 11th grade students. De Smet et al. (2019) focus on language attitudes in terms of perceived easiness and attractiveness of the TL and motivation in terms of expectancy for success, task value and cost, again using questionnaire data. Results point to more positive attitudes and higher motivation in CLIL compared to non-CLIL and in English compared to Dutch, mainly in the case of the secondary students. In addition, effect sizes suggest that the TL plays a more crucial role in language attitudes and motivation than the educational approach.

As can be seen, there is a great deal of variation across studies investigating motivation in a CLIL environment. While most studies have found CLIL students to have higher motivation than their non-CLIL peers, this finding has not necessarily always been statistically significant. Thus, while some studies (e.g., Lasagabaster & López Beloqui, 2015; Verspoor et al., 2015; Pfenninger, 2016; De Smet et al., 2018, 2019) point to a clear CLIL advantage, others (e.g., Heras & Lasabagaster, 2015; Rumlich, 2016, 2017; Lasabagaster, 2017) indicate no statistically significant difference between the two groups or find non-CLIL students to have higher motivation than their CLIL peers (e.g., Fernández Fontecha & Canga Alonso, 2014). These disparities are somewhat to be expected, given the very different methods used, diverse learning environments, and differences in students being compared (age, hours of exposure, level of programme intensity, requirements for programme, number of students, etc.) in the research to date.

3.5.3.2. Stakeholders' beliefs in CLIL. Though far less common than the research discussed above, other studies on motivation in CLIL have concentrated on stakeholders' beliefs, as summarised in Table 3.9 and discussed below.

Table 3.9

Motivation Research on Stakeholders' Beliefs in CLIL

Authors	Participants	Focus	Main Findings
Pladevall Ballester (2015)	197 students (5 th grade) 5 teachers 159 parents	Perceptions of CLIL in stakeholders, development of language learning motivation	General satisfaction among students, while teachers and parents express concerns regarding language level and acquisition of content.
Somer & Llinares (2018)	157 students (7 th grade)	Motivation towards both content and language as an integrated construct	Higher motivation (enjoyment and perceptions of its usefulness in the future) in students in a high-intensity CLIL track.
Ráez Padilla (2018)	237 parents	Perspectives on CLIL implementation	Generally positive view of CLIL, though issues were expressed regarding offering their children extramural support at home.
Campillo et al. (2019)	129 teachers	Perceptions of aspects of CLIL in science and social science in primary education	Positive view of native-speaker collaboration in the classroom by more experienced teachers, as it helps increase motivation and interculturality in the CLIL classroom.
San Isidro & Lasagabaster (2020)	44 students (9 th to 10 th grade) 44 parents	Attitudes to language learning and CLIL	Students and parents developed long-term positive attitudes and motivation toward language learning. CLIL students and their parents did so to a greater extent.

Table 3.9 (continued)

Barrios & Milla Lara (2020)	544 students (6 th and 10 th grade) 92 teachers 237 parents	Perspectives on methodology, materials and resources and assessment procedures	Key strengths include innovative pedagogical practices, whereas a key weakness was inadequate teaching training. Between-group comparisons revealed differences in groups' perceptions.
Barrios & Acosta-Manzano (2020)	524 students (aged 9 to 13 years)	Satisfaction with CLIL and perceived CLIL linguistic difficulty	High level of satisfaction which is related to linguistic difficulty; socioeconomic factors also appear to affect these factors.

Firstly, Pladevall Ballester (2015) analysed stakeholders' perceptions of CLIL and the development of language learning motivation over a period of two years in Catalonia, a bilingual region in Spain. The study included 197 students (154 of whom came from three schools implementing CLIL in science classes and 43 of whom came from two schools implementing CLIL in arts and crafts), 5 teachers and 159 parents. The participants' perceptions of CLIL were investigated by means of opinion questionnaires and interviews. Results found that children, except for low achievers, reported general satisfaction; teachers had concerns regarding students' level of English, lack of materials and support, and lack of competency in the content subject; and parents viewed CLIL as the only solution to their children's low English level, but also feared it would hinder their L1 and content knowledge.

Somers and Llinares (2018) investigated motivation not only towards the TL, but towards CLIL as an integrated construct. Participants included 157 7th grade students divided into those taking a high- ($n = 134$) or low- ($n = 23$) intensity track and were assessed by means of a 61-item questionnaire. Results indicated that track intensity had a significant impact on both the extent to which students enjoyed CLIL and their perceptions of its usefulness in their futures, with students in the high-intensity track expressing greater CLIL motivation.

Ráez Padilla (2018) focused on 237 parents, 152 of whom represented secondary

education and 85 of whom represented primary education. The instrument consisted in a self-administered questionnaire which included questions on biographical information and 40 opinion or value questions. Results indicated that parents overall see CLIL to exert positive effects on their children's language level and motivation, interest, and participation in the bilingual classroom. Issues, however, arose regarding parents' own self-appraisal when offering their children extramural support at home.

Campill, Sánchez and Miralles (2019) analysed 129 teachers' perceptions of aspects of CLIL in science and social science in primary education by means of a questionnaire. Of particular interest concerning motivation was that the collaboration of native speakers in the classroom was found to be more highly viewed by the more experienced teachers, given that the native speakers helped to increase motivation and interculturality in the CLIL classroom.

San Isidro and Lasagabaster (2020) offer a two-year longitudinal study involving 20 CLIL and 24 non-CLIL students in 9th to 10th grade and their families, in the bilingual region of Galicia, Spain. Stakeholders' attitudes and motivation towards learning a language in a multilingual CLIL environment were assessed by means of an adapted version of Gardner's AMTB. Results found that while both groups and their parents developed long-term positive attitudes and motivation toward language learning, the CLIL students and their parents did so to a greater extent.

Barrios and Milla Lara (2020) provide research on the perspectives of 544 students, 92 teachers and 237 parents in Andalusia, Spain. These perspectives, which dealt with methodology, materials and resources and assessment procedures, were measured by means of questionnaires and focus groups. Results showed that a key strength of the programme was its use of innovative pedagogical practices, whereas a key weakness was inadequate teaching training. Between-group comparisons also revealed differences in groups' perceptions, leading the researchers to highlight the vital need to increase awareness around how different stakeholders interpret the CLIL experience.

Finally, Barrios and Acosta-Manzano (2020) analysed 524 primary students' satisfaction with CLIL and perception of CLIL linguistic difficulty, determining the relationship between the two and between them and individual and social factors. In general, participants report high satisfaction with CLIL and do not find it linguistically challenging, though 20% did experience mild to severe difficulties in this regard. In addition, a moderate to strong correlation was found between satisfaction and linguistic difficulty, which were both related to mother's L1, mother's level of education,

availability to help with homework, and relatives' use of English in the workplace.

In summary, stakeholders' views of CLIL have in general been found to be quite positive, and view the approach as motivational, particularly in more intensive tracks (Somer & Llinares, 2018). Parents and teachers, however, have expressed concerns over a number of issues such as language level and acquisition of content (Pladevall Ballester (2015), how parents can offer extramural support in a language they are unfamiliar with (Ráez Padilla, 2018), and inadequate teacher training (Barrios & Milla Lara, 2020).

3.6. Research on Motivation and Gender

According to Ryan (2009, p. 135), there is a common assumption that FL learning is something of a “feminine terrain”, an idea which Lasagabaster (2016) attributes to the empirical evidence which has found females to be more motivated than their male peers. For example, early research such as Dörnyei et al.'s (2006) large-scale study on Hungarian school children, found that females consistently scored higher on various motivational variables than male students. Research has also shown that reasons for learning the L2 vary between different genders: males tend to be more instrumentally motivated, for example, with views to get a job or a better mark, while females tend to be more integratively motivated, for example, to integrate within the foreign culture (Ludwig, 1983; Powell & Littlewood, 1983; Mori & Gobel, 2006; Ghazvini & Khajepour, 2011). As a result of such findings, researchers have highlighted the vital role of gender in attitudinal and motivation research (Henry, 2009).

While the topic of gender will be dealt with specifically in more detail in Chapter 4, this section provides an overview of some of the key research investigating the relationship between motivation and gender, focusing in particular on research which has also incorporated other key issues, such as vocabulary, CLIL or L3 motivation. This research is summarised in Table 3.10 and discussed in turn below.

Table 3.10*Research on Gender and Motivation*

Authors	Participants	Focus	Main Findings
Merisuo-Storm (2006)	4 th grade (67 male, 78 female)	Attitudes towards FL learning, reading, and writing in CLIL; gender	Girls report more positive attitudes to both reading and writing.
Lasagabaster & Sierra (2009)	9 th + 10 th grade 172 CLIL 115 non-CLIL (40.3% male, 59.7% female)	Language attitudes; gender and social class	Females had statistically significant higher motivation than males in both teaching contexts.
Fernández Fontecha (2010)	8 th grade (139 male, 111 female)	LA, gender and motivation	Female students had higher motivation than male students; positive correlations between motivation and achievement, with a stronger correlation for female students.
Fernández Fontecha (2014a)	5 th grade (30 male, 25 female)	Receptive vocabulary, gender and motivation	No statistically significant difference between genders; positive correlation between male students' intrinsic motivation and receptive vocabulary.
Fernández Fontecha (2014b)	6 th grade (38 male, 28 female)		Female students were found to be more intrinsically motivated than male students. No correlation between receptive vocabulary and intrinsic and extrinsic motivation.

Table 3.10 (continued)

Fernández Fontecha & Canga Alonso (2014)	4 th grade (35 male, 27 female)	Motivation, gender and CLIL	No statistically significant difference between genders, though results varied depending on the teaching context.
Sylvén & Thompson (2015)	10 th grade (58 male, 119 female)	Motivation, gender and CLIL	Statistically significant differences between males and females in general; females report higher cultural interest, interest in FLs, international empathy, travel orientation and intended learning effort; females also had higher English anxiety and lack of L2 self-confidence.
Heras & Lasagabaster (2015)	10 th grade 25 CLIL (13 male, 12 female) 21 non-CLIL (11 male, 10 female)	Affective factors, vocabulary and gender	Males had higher Ought-to Self than females in CLIL; females had higher <i>Ideal L2 Self</i> than males in non-CLIL.
Lasagabaster (2016)	189 university students (88 male, 99 female)	L2MSS, gender, EMI	Statistically significant gender differences in the scales criterion measures, instrumentality-promotion, attitudes to EMI, attitudes to L2 community, and integrativeness; however, very small effect size.
Calafato & Tang (2019)	13-14 years old (24 male, 49 female)	L3 Motivational self-concept and gender	Statistically significant differences between the genders, particularly in the case of the L3 as opposed to L2 English.

Table 3.10 (continued)

Gallardo-del-Puerto & Blanco-Suárez (2021)	4 th to 6 th grade 124 CLIL (66 male, 58 female) 128 non-CLIL (60 male, 68 female)	Motivation, gender and CLIL	Gender differences were dependent on the teaching context: non-CLIL boys reported lower overall and intrinsic motivation than non-CLIL girls, but such no difference in CLIL.
--	--	-----------------------------	---

In a Spanish context, language learning motivation has often been assessed by means of Gardner’s (1985) AMTB, alongside vocabulary and/or CLIL education. For example, in addition to comparing participants’ LA or receptive vocabulary, Fernández Fontecha (2010, 2014b, 2014c) compared male and female learners in terms of motivation. In Fernández Fontecha (2010), which compared 139 male and 111 female 8th grade students, results indicated that there was a statistically significant difference between the participants’ motivation, with female students reporting higher motivation. A statistically significant correlation was also found between motivation and LA, which additionally was stronger for girls than for boys. Fernández Fontecha (2014a) compared 30 male and 25 female 5th grade students. Results indicated that there was no statistically significant difference between the male and female students in terms of motivation. However, a positive correlation was identified between male students’ intrinsic motivation and their receptive vocabulary. Fernández Fontecha (2014b), which compared 38 male and 28 female 6th grade students, did however find differences in terms of motivation: a statistically significant difference was found between the two groups in terms of intrinsic motivation, with girls reporting higher intrinsic motivation than their male peers. However, there was no statistically significant difference in extrinsic motivation. In addition, no correlation was found between receptive vocabulary size and intrinsic and extrinsic motivation. Fernández Fontecha and Canga Alonso (2014) compared the motivation of 62 4th grade male and female students in CLIL (17 male, 14 female) and non-CLIL settings (18 male, 13 female). Results indicated that gender differences varied depending on the teaching context. While non-CLIL girls were somewhat more motivated than non-CLIL boys, in a CLIL context the opposite was true: boys were somewhat more motivated than girls. These results, however, were not statistically significant. Using a seven-point semantic differential questionnaire based on Gardner (1985), Lasagabaster

and Sierra (2009) analysed the effect of gender on language attitudes in 172 CLIL and 115 non-CLIL students in 9th and 10th grade (40.3% male, 59.7% female). Results indicated that gender had a statistically significant effect on both CLIL and non-CLIL groups, with female students reporting higher motivation than male students. Heras and Lasagabaster (2015) investigated gender differences in terms of vocabulary and motivation in a CLIL context. Participants included 13 male and 12 female CLIL students and 11 male and 10 female non-CLIL students in 10th grade. With regards to motivation, results indicated that gender differences varied depending on the teaching context. CLIL males had a statistically significant higher Ought-to L2 Self than CLIL females, while non-CLIL females had a statistically significant higher Ideal L2 Self than non-CLIL males. Lasagabaster (2016) also compared male and female learners' motivation, this time using Taguchi et al.'s (2009) motivation questionnaire with EMI university students (88 male, 99 female). Results revealed statistically significant differences in five scales (criterion measures, instrumentality-promotion, attitudes to EMI, attitudes to L2 community, and integrativeness), with females in all cases producing higher means. However, given that the effect size is particularly small, the author points out that their magnitude is not meaningful. Finally, Gallardo-del-Puerto and Blanco-Suárez (2021) investigated gender and motivation in CLIL and non-CLIL primary level students. Participants included 252 students, 124 of which were CLIL (66 male, 58 female) and 128 of which were non-CLIL (60 male, 68 female) in 4th to 6th grade. The instrument used was an adapted version of Gardner's AMTB, designed specifically for young learners and consisting of a total of 34 items measuring intrinsic and extrinsic motivation. Results revealed that although non-CLIL males had significantly lower motivation than their non-CLIL female peers in overall motivation and intrinsic motivation, the same was not found in a CLIL setting. The authors thus highlight the potential levelling effect of this context, which may provide a more egalitarian educational setting in terms of language motivation and gender.

Elsewhere, notable contributions on gender and motivation, discussed previously and which also take multilingual education into consideration, have come from Merisuo-Storm (2006) in Finland, Sylvén and Thompson (2015) in Sweden, and Calafato and Tang (2019) in the United Arab Emirates. Regarding attitudes, Merisuo-Storm (2006) carried out a study with 145 4th grade Finnish learners of English (67 boys and 78 girls) aged between 10 and 11 years old. Participants' attitudes towards reading and writing were analysed using instruments based on McKenna and Kear's (1999) Elementary Reading

Attitude Survey and Kear et al.'s (2000) Writing Attitude Survey. Results showed firstly that there was a statistically significant correlation between attitudes to reading and writing, with those showing more interest in reading also showing more interest in writing. However, male students were found to enjoy writing less. In addition, female students' attitudes were significantly more positive towards both reading and writing. Differences were also found in the types of texts which were more appealing to each gender, with boys showing more interest in comics and humorous books while girls tended to prefer adventure books. Suggestions are made regarding the importance of meaningful purpose and communicative function in writing tasks in order to interest male students. In addition, boys' and girls' attitudes differed depending on the language learning context (see Merisuo-Storm, 2007, in [Section 4.3.3](#) for a discussion of this research with regards to CLIL). Sylvén and Thompson (2015) investigated the motivation of 177 10th grade students (58 male, 119 female) across three different schools using Ryan's (2009) MFQ on the L2MSS. Results revealed statistically significant differences between males and females in general, with female students reporting higher cultural interest, interest in FLs, international empathy, travel orientation and intended learning effort; however, it was also found that they had higher English anxiety and lack of L2 self-confidence. Finally, as discussed previously, Calafato and Tang (2019) investigated the L3 motivational self-concepts of 73 Arab teenagers who studied in English-medium schools. Upon comparing English and the students' L3 (French, Spanish, German, Japanese, Dutch, Italian, Russian, Turkish or Korean), results revealed statistically significant differences between male and female learners in motivational intensity, particularly in the case of L3 self-concepts. While female participants' Ideal L3 Self displayed much greater motivational intensity, the Ought-to L3 self did not differ to the same degree. Male students, on the other hand, generally exhibited higher motivational intensity in terms of their ideal English selves. Both genders were found to be more extrinsically motivated to learn the TL.

In conclusion, the majority of the studies reviewed above have found statistically significant differences between genders in terms of motivation, with female participants reporting higher motivation or positive attitudes (e.g., Fernández Fontecha, 2010; Fernández Fontecha, 2014c, for intrinsic motivation; Lasagabaster, 2016; Merisuo-Storm, 2006; Sylvén & Thompson, 2015; Calafato & Tang, 2019). However, results from some studies have shown no statistically significant differences, despite having very similar populations to some of those in the above studies (e.g., Fernández Fontecha, 2014b, 2014c for extrinsic motivation). Others, such as Sylvén and Thompson (2015) have also found

that in addition to reporting higher motivation, female students also report higher language learning anxiety and lower L2 self-confidence. Studies which also investigate vocabulary have also found some interesting gender-dependent correlations between motivation and language achievement. For example, Fernández Fontecha (2010) found a statistically significant correlation between motivation and LA in 8th grade students, which was stronger for females than males. Fernández Fontecha (2014a), on the other hand, found a positive correlation between male students' intrinsic motivation and receptive vocabulary in 5th grade. However, no such relationship was found in Fernández Fontecha (2014b) with regards to 6th grade students, where no correlation was found between receptive vocabulary and intrinsic or extrinsic motivation. Others, such as Fernández Fontecha and Canga Alonso (2014) and Gallardo-del-Puerto and Blanco-Suárez (2021), have found that gender-based differences may be context-dependent. In the former study, female students had higher motivation in a non-CLIL context, but boys had higher motivation in a CLIL context, though the difference was not statistically significant. In the latter study, boys had statistically significant lower overall motivation and intrinsic motivation in a non-CLIL context, but no such differences in motivation were observed in a CLIL context. Such findings highlight the importance of taking into consideration not only motivation and gender in isolation, but also other individual and contextual variables so as to determine the interrelated nature of these various factors.

Chapter 4: Gender

This chapter opens with a brief introduction to the construct of gender with regards to language acquisition in general. It then offers an overview of gender specifically with regards to language learning, highlighting the key similarities and differences in gender and SLA. Finally, it focuses in particular on the most pertinent research that has been carried out on gender specifically with regards to three other key issues in this thesis: vocabulary, motivation and CLIL.

4.1. Gender and Language Acquisition

As is highlighted by Talbot (2019), though the terms sex and gender are often used interchangeably, second-wave feminism, and in particular Oakley (1972), highlighted a key distinction between the two terms: while sex is biologically founded, gender is socially constructed. This entails that while sex is generally understood as a dimorphic division of a species according to bodily attributes, gender is “a construct shaped by historical, cultural, social, and interactional factors” (Ehrlick, 1997, p. 424) and is essentially learned, with individuals acquiring features which are perceived to be masculine or feminine¹ (Talbot, 2019). In keeping with this distinction, and with regards to language acquisition, gender may be seen both as a biological as well as a social factor (Saville-Troike & Barto, 2017). Gender studies has consequently been marked by two competing theories which attempt to explain language differences between males and females, namely the biological theory and the sociological theory (Bell, McCarthy & McNamara, 2006). The former sees gender roles as static and contextually independent, paying little attention to language individualisation (Coates & Johnson, 2001). The latter, on the other hand, views gender roles as fluid and contextually situated and maintains that men and women are by no means restricted to one specific language style, but rather adopt different styles according to the social context of their interactions (Leaper & Smith, 2004).

As far back as 1954, a review of the literature on sex and language acquisition led McCarthy (1954, p. 580) to conclude that there was “convincing proof that a real sex difference in language development exists in favour of girls”. However, there has since been considerable debate on the topic. Researchers such as Halpern (2000) have noted that these differences have not always been so evident, and a much more recent,

¹ Henceforth, the term gender will be used throughout this thesis, unless specifically referring to research investigating sex as a biological construct.

systematic review on sex differences in childhood language and brain development carried out by Etchell et al. (2018) concluded that sex differences in language still remain unclear and noted that the opposing views in the field are in keeping with inconsistent findings in research into factors affecting sexual dimorphism in language development. Hartshorne and Ullman (2006) have also attributed these inconsistencies to research's lack of focus on the sub-components of language and the exploratory rather than hypothesis-driven nature of the research.

Yet despite this lack of clarity in empirical evidence, there remains a prevalent assumption that boys lag behind girls in language development (Barbu et al., 2015). According to Oxford (1989), this gender difference has been commonly associated with women's greater social orientation, stronger verbal skills and greater conformity to norms. Regarding specific language areas, one notable large-scale study was carried out by Eriksson et al. (2012) with almost 14,000 children in 10 non-English language communities in Europe, investigating emerging language skills. Girls were found to outperform boys in early communicative gestures, productive vocabulary, and in combining words, a difference which increased with age. More variation was also found in male learners. Barbu et al. (2015) have also highlighted a female advantage in the first 30 months of life with regards to communicative gestures, early vocabulary growth, morphosyntactic growth, vocabulary size and syntactic complexity. While these authors again stress the lack of systematicity across studies in these language skills, they highlight that more consistent findings have been reported indicating a female advantage regarding vocabulary production. Thus, while research is by no means clear, females are often thought to outperform males in L1 acquisition, particularly with regard to vocabulary skills.

4.2. Gender in Second Language Acquisition

Turning to gender differences in SLA, as in the case of L1 acquisition, a distinction has again been made between sex and gender (Ellis, 1994). However, as Jiménez Catalán (2005) astutely points out, sex and gender may be seen as two sides of the same coin: both variables influence the use of language by men and women, but it is unclear which one is dominant in a particular context. This issue was taken up by Ellis (1994), who maintains that given gender's interaction with other factors such as age, ethnicity and social class, females will not always necessarily outperform males: Asian men in Britain may have outperformed Asian women simply because their job entailed

higher L2 input, while women may have been restricted to the home context. Saville-Troike and Barto (2017) have also highlighted that, despite the widespread belief that women outperform men in language learning, this is likely first and foremost a social construct, whereby outcomes reflect cultural and sociopsychological constraints and influences. Thus, SLA research into sex and gender must inevitably take into consideration not only the neurological differences, but also the cognitive, affective, social and educational variables, as well as the interaction of these various factors (López Rúa, 2006).

Regarding empirical evidence on gender differences in SLA, van der Slik, van Hout and Schepens (2015) highlight the absence of research, which they suggest is due to a female advantage simply being taken for granted. This assumption may be somewhat well-founded as, although the little research which has been carried out has provided mixed results, there does again seem to be a general trend indicating a female advantage (Bowden, Sanz & Stafford, 2005). This research has generally focused on achievement (how well males and females perform in different language areas), strategies (what learning strategies are employed) and affective factors (learner's attitudes, motivation and self-confidence). Each of these three areas will now be discussed in turn, indicating some of the main findings and trends in SLA research.

Firstly, regarding achievement, from a biological point of view, some research has suggested that differences may be related to hormonal variables. As is explained by Saville-Troike and Barto (2017, p. 90), "higher androgen level correlates with better automatized skills, and high estrogen with better semantic/interpretive skills". Estrogen has also been suggested to modulate verbal memory, which also affects memory for complex forms in the L1 (Bowden et al., 2005). Regarding processing, Saville-Troike and Barto (2017) note that, although evidence from language acquisition research is mixed, there are some important findings which may be relevant to SLA. For example, findings by Kimura (1992) have found women to outperform men in some verbal fluency tests and shown that women's brains may not be as asymmetrically organised as men's for speech, while Halpern (2000) points to differences in mental representations in the lexicon as opposed to grammar. This research suggests that while women appear to be better at memorising complex forms, men may be better at computing compositional rules. According to Bowden et al. (2005), given the findings which support the belief that women excel at verbal memory, Ullman (2004) has suggested a female advantage in lexical abilities and memorisation of previously encountered complex forms in the

declarative system. Men, on the other hand, are more likely to rule-compute these forms in the procedural system in real time. Regarding general language achievement, López Rúa (2006) notes that girls generally outperform boys in overall language achievement, and FLs in particular, though some disagreement is highlighted regarding performance in individual skills such as listening. In addition, girls tend to excel in particular at verbal skills tasks while boys do better in visual and spatial ability tasks. Pae (2004) analysed reading comprehension in 14,000 pre-college Korean EFL learners and measured the gender effect by means of a Differential Item Functioning methodology. Results indicated that females outperformed males, but more importantly that while items classified as Mood/Impression/Tone were easier for females, those classified as Logical Inference were easier for males. These results highlight the importance of item type when administering reading tests to male and female L2 learners. As for adult learners, van der Slik et al. (2015) carried out a large-scale study with almost 30,000 adult learners of Dutch, who came from a wide range of different countries and had different L1s. Results found females to consistently outperform males in productive skills (speaking and writing), while no gender gap was found for receptive skills (listening and reading). In addition, a general gender by education effect was found for all four language skills. Regarding the different findings for productive and receptive skills, the authors also highlight the issue that these different modalities are generally assessed using different formats. They emphasise findings by Walstad and Robson (1997) which indicate that females tend to perform poorly on multiple-choice tests (often used when measuring receptive skills), while findings by Lumsden and Scott (1987) suggest that they do better on open format tests (often used when measuring productive skills). This may in part explain the lack of clarity across different studies, which may inevitably use different testing formats which could be better suited to males or females.

Secondly, regarding language learning strategies (LLS), research has pointed to a significant association between gender and choice of LLS at a superficial level (Liyanage & Bartlett, 2012). As discussed by Bowden et al. (2005), early research by Politzer (1983), Gass and Varonis (1986), and Ehrman and Oxford (1989) found that female learners use more LLSs than male learners. Similar results were found more recently by Aslan (2009), who additionally found a significant relationship between gender, LLSs and achievement in English. In addition, Oxford, Nyikos and Ehrman (1988), following a review of several studies, found that not only do females employ more LLSs than males, but they also do so more often. Regarding the differences in the types of LLSs that are employed, Bacon

(1992) found that men tended to use more translation strategies, while women monitored their comprehension to a greater extent. Regarding vocabulary learning strategies, Jiménez Catalán (2003) again found higher use in females, and found that while they used more formal rule, input elicitation, rehearsal, and planning LLSs, males tended to use image vocabulary learning strategies (see [Section 4.3.1](#) for further discussion).

Finally, regarding affective factors, Ellis (1994) has pointed again to a female advantage in terms of affective factors such as motivation and attitudes towards L2 speakers. This is also highlighted in López Rúa (2006), who maintains that girls have been found to have more positive attitudes towards language learning, which leads to higher motivation and in turn higher language proficiency. This author also suggests that girls' higher aptitude may enhance their confidence, which then contributes to their achievement. Summarising some of the key evidence in gender and SLA research, López Rúa (2006) also notes that more girls decide to study FLs and take language examinations than boys; that they tend to be more confident in their ability to master the language while boys are more self-deprecating; and that sex-stereotyping of jobs supports language learning as an accomplishment for girls, making them more likely to perceive languages as vocationally relevant. Given the importance of motivation in language gains, such findings are imperative in interpreting any differences in gender in SLA. Research that has investigated gender differences in motivation, which was addressed in detail in Section 3.6, will be summarised and discussed specifically in Section 4.3.2.

As can be seen, there have been some general indications in SLA research that a gender difference exists, often in favour of female learners. The following section will delve further into this research in order to address the studies which are most pertinent to the research at hand.

4.3. Research on Gender and Vocabulary, Motivation and CLIL

The following three sections provide a more detailed review of SLA research on gender that has been carried out over the past decade, homing in on three areas which are of central importance to this thesis: vocabulary, motivation and CLIL.

4.3.1. Gender and Vocabulary

As has been discussed above, vocabulary is an area in which research has shown a clear advantage for female learners, both in first and second language acquisition. Given the crucial role vocabulary plays in FL and L2 acquisition and its relation to language achievement (Sunderland, 2010), this implication is incredibly important as it implies that

girls may also be superior language learners. According to Agustín Llach and Fernández Fontecha (2014), the female advantage in vocabulary studies may in fact also be attributable to female students' greater language learning motivation. They also highlight Hernández Muñoz's (2010) suggestion that both cognitive and attitudinal differences play a part: lexical categorization differs in males and females, and in addition, girls may take the task at hand more seriously when being assessed. As is pointed out by Sunderland (2010), research on vocabulary and gender is, however, scarce and given the range of ages, levels, classroom activities and social contexts under analysis, unsurprisingly yields varied results. Research addressing gender and vocabulary in a Spanish context has largely focused on either LLSs or language gains in terms of receptive or productive vocabulary. With regard to that on productive vocabulary, an extremely important contribution has been the 2010 volume *Gender Perspectives on Vocabulary in Foreign and Second languages*, edited by Rosa María Jiménez Catalán, which provides an ample range of empirical studies on the topic, and includes research on adult, adolescent and young learners of English and Spanish in a classroom context. Particularly relevant contributions include those by Díez Prados (2010), Agustín Llach (2010), Fernández Fontecha (2010), Jiménez Catalán (2010) and Moreno Espinosa (2010). These studies and some of the other key research on gender and vocabulary is summarised in Table 4.1 and discussed below.

As shown, research investigating vocabulary and gender may be divided into that which investigates LLSs (Jiménez Catalán, 2003; Montero-Saiz Aja, 2021), receptive vocabulary size (Agustín Llach & Terrazas Gallego, 2012; Fernández Fontecha, 2014b; Fernández Fontecha, 2014c), both receptive and productive tests (Jiménez Catalán, 2010; de la Maya Retamar, 2016), productive vocabulary (Díez Prados, 2010; Agustín Llach, 2010; Moreno Espinosa, 2010; Montero-Saiz Aja, 2021), and LA (Jiménez Catalán & Ojeda Alba, 2009a; Fernández Fontecha, 2010; Agustín Llach & Fernández Fontecha, 2014; Jiménez Catalán & Canga Alonso, 2019).

Table 4.1*Research on Gender and Vocabulary in a Spanish Context*

Authors	Participants	Focus	Main Findings
Jiménez Catalán (2003)	Primary, secondary and university students and adults (279 male, 302 female)	Number and range of LLSs	Females used a significantly higher number of LLSs; differences in the types of strategies used by each gender.
Jiménez Catalán & Ojeda Alba (2009a)	6 th grade (105 male, 105 female)	LA	Statistically significant differences between genders: girls achieved higher means than boys in all fifteen prompts of the LAT.
Jiménez Catalán (2010)	6 th grade (105 male, 105 female)	Receptive and Productive Tests	Similarities in receptive tests but differences in productive tests.
Diéz Prados (2010)	University students (130 male, 2141 female)	Productive vocabulary: Lexis in writing	Men tended to have higher lexical variation; no difference was found between genders in lexical density; females produced more nominalizations and certain adverbs.
Agustín Llach (2010)	8 th grade (168 male, 130 female)	Productive vocabulary: Lexical Creations	High degree of similarity in the number and type of lexical creations.
Moreno Espinosa (2010)	4 th , 5 th and 6 th grade (124 male, 101 female)	Productive vocabulary: Word Associations	Similarities between genders, though girls in 6 th grade produced a higher number of tokens.

Table 4.1 (continued)

Fernández Fontecha (2010)	8 th grade (139 male, 111 female)	LA	Statistically significant differences between the two groups in favour of the female learners.
Agustín Llach & Terrazas Gallego (2012)	4 th to 9 th grade (94 male, 82 female)	Receptive vocabulary size	No size differences in mean scores; female learners showed higher vocabulary gains for first three intervals while at last two intervals they were higher for males.
Fernández Fontecha (2014a)	5 th grade (30 male, 25 female)	Receptive vocabulary size	No statistically significant difference between genders in receptive vocabulary.
Fernández Fontecha (2014b)	6 th grade (38 male, 28 female)	Receptive vocabulary size	No statistically significant difference between genders in receptive vocabulary.
Agustín Llach & Fernández Fontecha (2014)	6 th + 9 th grade (106 male, 84 female)	LA	Statistically significant differences between genders: girls produced a higher number of words; similarities across genders in most and least productive prompts.
de la Maya Retamar (2016)	8 th + 9 th grade (38 male, 43 female)	Receptive and productive vocabulary and LA in French, gender and motivation	Motivation correlated with productive vocabulary but not with receptive vocabulary or LA; no gender differences.

Table 4.1 (continued)

Jiménez Catalán & Canga Alonso (2019)	12 th grade (94 male, 171 female)	LA	No statistically significant difference between genders; some qualitative differences in the types of words produced.
Montero- Saiz Aja (2021)	12 th grade (20 male, 31 female)	LLSs and Productive Vocabulary	Females used LLSs significantly more than males; no statistically significant difference in productive vocabulary.

As previously mentioned, use of vocabulary learning strategies in males and females has been investigated by Jiménez Catalán (2003), who analysed the LLSs of 581 Spanish-speaking learners of Basque and English. Participants included 279 males and 302 females, who were aged between 11 and 56 and who studied English at primary, secondary or university level or, in the case of adult participants, took Basque courses in a language school in one of four levels: beginner, intermediate, advanced or proficiency. A questionnaire on vocabulary learning strategies was administered with the aim of determining both the number of and the range of strategies used. Regarding the former, results indicated a statistically significant difference in the number of strategies used by each gender, with females reporting a higher usage, though their mean was only slightly higher. Regarding the latter, while common patterns emerged, differences were found with regard to a number of strategies. For example, while females used more formal rule, input elicitation, rehearsal, and planning vocabulary learning strategies, males used more image vocabulary learning strategies. While the author highlights caution in this interpretation, given that data is derived from learners' own perceptions of learning behaviours rather than observing them directly, the results offer interesting possibilities regarding the vocabulary language learning styles of male and female learners. However, it is also interesting to note that male learners' higher use of image vocabulary learning strategies is consistent with the language achievement gains discussed by López Rua (2006), mentioned above, which maintains that male learners tend to be better at visual tasks. Montero-Saiz Aja (2021) also investigated gender-based differences in LLSs.

Participants, 51 12th grade EFL learners (20 male and 31 female), completed both the Strategy Inventory for Language Learning questionnaire and the PVLТ. Results of the former again indicated differences in the use of LLSs between genders: females used LLSs significantly more than their male peers. In addition, there was a statistically significant positive correlation between LLSs and productive vocabulary, indicating that those who used more LLSs also had a higher score on the PVLТ (see below).

With regards to receptive vocabulary and gender, notable contributions come from Agustín Llach and Terrazas Gallego (2012) and Fernández Fontecha (2014a, 2014b). Firstly, Agustín Llach and Terrazas Gallego (2012) analysed the receptive vocabulary of 176 students (94 male and 82 female) using the 2000-word frequency band from the receptive version of the VLT. Students were assessed over six consecutive years, starting when the students were in 4th grade of primary school and finally when they were in 9th grade of secondary school, thus providing longitudinal data. Results indicated that all participants increased their receptive vocabulary across the six years, with highly significant differences from one year to the next. Although females' average scores were higher at each testing period, no significant differences were found in the mean scores of male and female learners. However, differences were observed in the rate of the gains at different points: while girls had higher vocabulary gains at the first three intervals (4th to 6th grade), boys' gains were higher at the last two intervals (8th and 9th grade). Fernández Fontecha (2014a, 2014b) also investigated receptive vocabulary using the 2K VLT, focusing respectively on 55 5th grade (30 male, 25 female) and 66 6th grade (38 male, 28 female) Spanish EFL learners. Results from both studies again revealed no statistically significant difference between genders in terms of receptive vocabulary size.

Combining research into both receptive and productive vocabulary, Jiménez Catalán (2010) investigated the effect of gender in receptive and productive vocabulary tests. Participants included 210 Spanish EFL learners in 6th grade (105 boys, 105 girls). Receptive vocabulary was assessed by means of the 1K and 2K VLT, while productive vocabulary was assessed using a writing composition test and a cue word test. Results revealed statistically significant differences between the receptive tests (1K vs. 2K VLT) and between the productive tests (composition vs. cue words). Regarding gender, no significant differences were found between the two groups in either of the receptive tests, which is consistent with the results by Agustín Llach and Terrazas Gallego (2012) and Fernández Fontecha (2014a, 2014b) above. However, upon comparing them in the productive tests, girls were found to produce a significantly higher number of word types

and tokens than their male peers, indicating a higher level of lexical richness. Statistically significant positive correlations were also found between performance in the receptive and productive tests, indicating that students who performed well in one test also performed well in the others. As discussed in Section 2.5.3, de la Maya Retamar (2016) also analysed gender differences in both the receptive and productive vocabulary of Spanish students' French in 8th (16 male, 22 female) and 9th grade (17 male, 26 female). However, no differences were found between the male and female groups in any of the vocabulary measures (X-Lex test, Lex30 and LAT).

Focusing on productive vocabulary, Diéz Prados (2010) provided an analysis of the lexicon used in written compositions found in two corpora: the International *Corpus of Learner English* (intermediate-advanced level university students) and part of the MAD Corpus (American native-English university students). A total of 2,271 texts were analysed (2,111 from female non-native speakers, 115 from male non-native speakers, and 45 from native speakers) in terms of lexical variation (type-token ratio; TTR), lexical density, presence of grammatical metaphor, and lexical resources such as certainty and doubt adverbs. Results indicated significant differences in terms of lexical variation before error-removal in the case of female and male non-native speakers overall, and female and male non-native speakers with German as an L1, whereby men displayed higher lexical variation than women. This difference was notably not found in subgroups whose L1 was a Romance language. Following error-removal, significant differences were also found in favour of the male groups in the Spanish, French and Dutch populations. No differences were found between genders in lexical density. Finally, females produced more nominalizations and doubt adverbs than males, who used more boosting than hedging adverbs. Agustín Llach (2010) explored the role of lexical creations in 298 Spanish EFL learners (168 male, 130 female) in 8th grade of secondary school by means of a written composition (a letter to a host family). Texts were first analysed in terms of lexical inventions, which were subdivided into intralingual inventions, or “word coinages”, and interlingual inventions, which included “foreignizing” and “literal translations”. Results revealed a high degree of similarity in the number and type of lexical creations by male and female learners: no statistically significant differences were found between genders in the number of lexical inventions or the order of frequency in which they appear. Moreno Espinosa (2010) analysed word associations in 225 Spanish EFL learners in 4th, 5th, 6th grade of primary school (124 male, 101 female). Participants were assessed by means of the Lex30, in order to elicit their L2 word

association responses and thus determine the number of types and tokens recalled, the characteristics of the responses, and the part of speech of associations. Results firstly showed an increase in vocabulary across the three years, with mean types and tokens increasingly progressively. No statistically significant differences were found between the two groups in terms of these measures, with the exception of the number of tokens in 6th grade. No clear difference was revealed in terms of categories or the type of association, though older male students changed from a preference towards syntagmatic responses to paradigmatic associations. Finally, as mentioned above, in addition to investigating LLSs, Montero-Saiz Aja (2021) also compared 12th grade students' productive vocabulary by means of the PVL. Despite the differences which were found regarding LLSs, no statistically significant difference was found between genders in productive vocabulary. However, a statistically significant positive correlation was found between the participants' LLSs and their productive vocabulary, with a higher number of LLSs relating to a higher score on the PVL, though the correlation was weak.

Finally, as discussed in Section 2.5.3, a number of researchers have investigated gender differences with regards to LA. Jiménez Catalán and Ojeda Alba (2009a) analysed the LA of 210 young learners in 6th grade of primary school (105 male, 105 female) using 15 different prompts. Results revealed statistically significant differences in the average number of words produced by each group: girls achieved higher means than boys in all fifteen prompts of the LAT. Fernández Fontecha (2010) analysed gender differences in a cohort of 250 8th grade Spanish EFL learners (139 male, 111 female) in terms of LA and motivation (see [Section 4.3.2](#) for findings of the latter). Results indicated that there was a statistically significant difference between the two groups in favour of the female learners, who again produced a higher number of words. Agustín Llach and Fernández Fontecha (2014) addressed the same issue when comparing young learners and teenage learners. Participants included 190 Spanish EFL learners, 106 male and 84 females, whose LA was assessed using nine prompts at two different points: first while the learners were in 6th grade and then again three years later while they were in 9th grade. Results indicated a statistically significant difference between male and female learners at both levels, with girls again producing a higher number of words than boys, although the significance values were found to decrease as the learners grew older. This suggests that gender differences in LA may also be intertwined with other individual differences such as age. In addition, the results indicated that at both levels, males and females showed similarity in the most and least cued responses of different semantic fields: prompts such

as *Animals, Food and Drink* and *School* were the most productive, while those such as *Countryside, Transport* and *Professions* were the least productive for both genders and in both 6th and 9th grade. Lastly, Jiménez Catalán and Canga Alonso (2019) assessed the LA of 265 12th grade Spanish EFL learners (94 male, 171 female). In contrast to the previous studies, the results revealed that there was no statistically significant difference between the two groups in the number of words produced. These findings indicate a potential difference in 12th grade male and female students as compared to younger learners. Closer inspection of the actual words produced did, however, indicate some qualitative differences. For example, regarding the prompt *Professions*, words such as “engineer”, “politician” and “professor” were used only by male participants while “singer”, “musician” and “shop assistant” were used only by female participants.

In conclusion, regarding vocabulary and gender in a Spanish context, a number of key trends have been observed. In terms of receptive vocabulary, no significant differences between genders have been observed (Jiménez Catalán, 2010; Agustín Llach & Terrazas Gallego, 2012; Fernández Fontecha, 2014b; de la Maya Retamar, 2016). Findings for productive vocabulary analysed by means of written compositions are somewhat less clear. While differences were observed in terms of written compositions and cue words in favour of females (Jiménez Catalán, 2010) and lexical variation in favour of males (Diéz Prados, 2010), no differences were found in terms of lexical density (Diéz Prados, 2010), lexical creations (Agustín Llach, 2010), word associations (Moreno Espinosa, 2010), the Lex30 (de la Maya Retamar, 2016) or the PVLТ (Montero-Saiz Aja, 2021). On the other hand, statistically significant differences have been found between females and males in terms of the use of LLSs (Jiménez Catalán, 2003; Montero-Saiz Aja, 2021), and very clearly so in terms of English LA from 6th to 9th grade (Jiménez Catalán & Ojeda Alba, 2009a; Fernández Fontecha, 2010; Agustín Llach & Fernández-Fontecha, 2014), with female students using more LLSs and also producing more words in the LAT. On the other hand, in the limited research carried out on French LA in 8th and 9th grade, no statistically significant differences were found between male and female students (de la Maya Retamar, 2016). In addition, when analysing the English LA of older students in 12th grade, Jiménez Catalán and Canga Alonso (2019) found no such difference between male and female learners, indicating a potential age-related effect. There is thus a clear need for research to address the issue of gender differences in the LA of Spanish secondary students in grades that have not yet been addressed, namely 10th and 11th grade. This would enable us to determine whether there is indeed a clear difference between

genders in earlier levels which progressively decreases towards the end of secondary education.

4.3.2. Gender and Motivation

As detailed in Chapter 3, female learners have by and large been found to be more motivated than their male peers, a fact which has greatly influenced the view of FL learning as predominantly a female affair (Lasagabaster, 2016). This section offers a brief summary of some of this research, specifically that which has been outlined and discussed in detail in Section 3.6, so as to highlight the key research and findings related to the topic of gender and motivation. These studies have predominantly been carried out in a Spanish context and have all investigated gender and motivation alongside other key areas of interest, including vocabulary, CLIL and L3 acquisition.

In general, research into gender and motivation has indicated that female students generally report higher motivation or positive attitudes (e.g., Fernández Fontecha, 2010; Fernández Fontecha, 2014c, for intrinsic motivation; Lasagabaster & Sierra, 2009; Lasagabaster, 2016; Merisuo-Storm, 2006; Sylvén & Thompson, 2015; Calafato & Tang, 2019). Exceptions to this were found in Fernández Fontecha (2014a, 2014b), which found no statistically significant differences in male and female students in general in the former, and in terms of extrinsic motivation in the latter. Furthermore, while studies such as such as Sylvén and Thompson (2015) have found that female students report higher motivation, they have also reported higher language learning anxiety and lower L2 self-confidence. Regarding the relationship between gender, motivation and vocabulary, findings have been somewhat mixed. While significant correlations were found between 8th grade students' LA and motivation for both genders (Fernández Fontecha, 2010) and between 6th grade male students' receptive vocabulary and intrinsic motivation (Fernández Fontecha, 2014c), no correlation was found between 5th grade students' receptive vocabulary and intrinsic or extrinsic motivation for males or females (Fernández Fontecha, 2014b). Other studies further revealed that gender-based differences may be context-dependent (e.g., Fernández Fontecha & Canga Alonso, 2014; Lasagabaster and Sierra, 2009; Heras & Lasagabaster, 2015; Gallardo-del-Puerto & Blanco-Suárez, 2021), with male and female CLIL and non-CLIL showing clear differences with regard to motivation. It is thus incredibly important to consider this contextual variable when discussing gender and motivation, as discussed in the following final section.

4.3.3. Gender and CLIL

As noted in the previous sections, SLA research has largely pointed to the significant role of gender in language learning, finding often that females, who appear to be more motivated, are usually better language learners and outperform their male peers (San Isidro, 2010). However, some research has indicated that these differences may experience a change outside of the typical FL classroom. Within a Canadian immersion context, for example, Baker and MacIntyre (2000) have reported a levelling effect of gender-based differences, finding statistically significant differences in male and female students' orientation only in the non-immersion group under investigation. This has led European researchers to believe the same may be found in a CLIL, with Lasagabaster (2008) suggesting a potential blurring effect of gender-difference in a CLIL context. The reason for this effect is explained by Marsh (2000), who notes that while CLIL classes have a positive effect on students' desire towards FL learning, this is particularly so among male students. This may be because, given the presumed gender bias which sees females as optimal language learners and may consequently alienate males, CLIL can provide an alternative approach to language learning, which may serve to reduce this exclusion. In addition, as suggested by Heras and Lasagabaster (2015), male students may compensate lower FL learning motivation with higher motivation towards the CLIL subject. This is in keeping with Dalton-Puffer's (2008) findings, which will be discussed in Section 5.1.2.1, which indicate that CLIL has the ability to significantly enhance the L2 skills of the many students who have an average level.

Several of the studies discussed in the previous sections have also investigated gender-based differences in CLIL, alongside vocabulary and/or motivation (e.g., Merisuo-Storm, 2007; Fernández Fontecha & Canga Alonso, 2014; Lasagabaster and Sierra, 2009; Heras & Lasagabaster, 2015; Gallardo-del-Puerto & Blanco-Suárez, 2021). These studies and others addressing gender in CLIL are summarised in Table 4.2 and are discussed in turn below.

Table 4.2*Research on Gender in CLIL*

Authors	Participants	Focus	Main Findings
Merisuo-Storm (2007)	4 th grade 70 CLIL 75 non-CLIL (67 male, 78 female)	Attitudes towards FL learning, reading and writing; gender	A statistically significant difference was found between boys' and girls' attitudes in the non-CLIL group, but not in the CLIL group.
Lasagabaster (2008)	10 th grade 113 CLIL (47 male, 66 female)	FL competence in CLIL; gender	No levelling effect of CLIL: female student were found to outperform their male peers in all language tests.
Lasagabaster & Sierra (2009)	9 th + 10 th grade 172 CLIL 115 non-CLIL (40.3% male, 59.7% female)	Language attitudes; gender and social class	CLIL students had more positive attitudes towards English than non-CLIL students; females had more positive attitudes than males in both CLIL and non-CLIL.
Fernández Fontecha & Canga Alonso (2014)	4 th grade 31 CLIL (17 male, 14 female) 31 non-CLIL (18 male, 13 female)	Motivation and gender	No statistically significant difference between CLIL and non-CLIL, though non-CLIL boys and girls were more motivated than CLIL boys and girls respectively.
Fernández Fontecha (2014a)	5 th grade 55 CLIL (30 male, 25 female)	Receptive vocabulary, gender and motivation	No statistically significant difference in terms of motivation; positive correlation between male students' intrinsic motivation and receptive vocabulary.

Table 4.2 (continued)

Fernández Fontecha (2014b)	6 th grade 66 CLIL (38 male, 28 female)	Receptive vocabulary, gender and motivation	Statistically significant difference between male and female CLIL students in intrinsic motivation; No correlation between receptive vocabulary and intrinsic and extrinsic motivation.
Heras & Lasagabaster (2015)	10 th grade 25 CLIL (13 male, 12 female) 21 non-CLIL (11 male, 10 female)	Affective factors, vocabulary and gender	Males had higher Ought-to Self than females in CLIL; Females had higher <i>Ideal L2 Self</i> than males in non-CLIL; CLIL had a positive effect on vocabulary for both genders.
Sylvén & Thompson (2015)	10 th grade 109 CLIL (29 male, 80 female) 68 non-CLIL (29 male, 39 female)	L1, motivation and gender	Statistically significant differences in L2 self-confidence and English anxiety depending on gender and learning context.
Gallardo-del-Puerto & Blanco-Suárez (2021)	4 th to 6 th grade 124 CLIL (66 male, 58 female) 128 non-CLIL (60 male, 68 female)	Motivation, gender and CLIL	Gender differences were dependent on the teaching context: non-CLIL boys had lower overall and intrinsic motivation than non-CLIL girls, but such no difference in CLIL.

Regarding young learners, when comparing the motivation of the 4th grade Finnish students in her study, Merisuo-Storm (2007) also compared CLIL and non-CLIL students

by means of two instruments based on McKenna and Kear's (1999) Elementary Reading Attitude Survey and Kear et al.'s (2000) Writing Attitude Survey. Results indicated that the difference in the attitudes was not as apparent in the bilingual group as in the monolingual group. In addition, while a statistically significant difference was found between boys' and girls' attitudes towards language learning in the non-CLIL group, the same was not found in the CLIL group. In Spain, Fernández Fontecha and Canga Alonso (2014) and Fernández Fontecha (2014a, 2014b), discussed above, assessed primary level CLIL students using Gardner's (1985) AMTB, with the first study also comparing these students to non-CLIL students. Fernández Fontecha and Canga Alonso (2014) thus assessed 17 male and 14 female CLIL students and 18 male and 13 female non-CLIL students in 4th grade. Results indicated that while girls were slightly more motivated than boys in the non-CLIL group, the boys were slightly more motivated than girls in the CLIL group. In addition, non-CLIL boys and non-CLIL girls were more motivated than CLIL boys and CLIL girls, respectively. However, in no cases were these results statistically significant. Fernández Fontecha's (2014a) participants included 55 5th grade CLIL students (30 male and 25 female). Results found no statistically significant difference between the two genders in terms of motivation. However, a positive correlation was found between male students' intrinsic motivation and their receptive vocabulary. Fernández Fontecha (2014b) carried out a study involving 66 CLIL learners (38 male and 28 female) in 6th grade, analysing the effect of gender on receptive vocabulary size and motivation. While results showed no statistically significant difference in the vocabulary size of the male and female students, they indicated that the female students were more intrinsically and extrinsically motivated than males, with a statistically significant difference only for intrinsic motivation.

Regarding teenagers, Lasagabaster and Sierra (2009) analysed the effect of gender on language attitudes in CLIL and non-CLIL students in 9th and 10th grade using a seven-point semantic differential questionnaire based on Gardner (1985). Findings indicated that the CLIL group had more positive attitudes towards English than the non-CLIL group, the results of which were statistically significant. Gender also had a significant effect on both CLIL and non-CLIL groups, with female students reporting more positive attitudes than male students in both cases. Lasagabaster (2008) compared 47 male and 66 female CLIL 10th grade students in terms of their speaking, writing, grammar, listening and overall English competence. Despite the expectation that CLIL would help to balance gender-related differences, female students were found to outperform their male peers in

all tests. These differences were statistically significant in all tests except for the speaking test. Heras and Lasagabaster (2015) also assessed 10th grade students, addressing vocabulary, motivation and gender in a CLIL context, and comparing 13 male and 12 female CLIL students and 11 male and 10 female non-CLIL students. Regarding motivation, results revealed that, in both CLIL and non-CLIL, females had higher means in instrumental motivation and the Ideal L2 Self, whereas males had higher means in Ought-to L2 Self, self-esteem in the PE CLIL class, and self-esteem in the FL class. In the CLIL group, only the difference in the Ought-to L2 Self was found to be statistically significant, while in the non-CLIL group the only difference which was statistically significant was the Ideal L2 Self. Regarding vocabulary, results showed that the CLIL approach had a positive effect on both male and female students' technical content-related vocabulary. In Sylvén and Thompson's (2015) study, 29 male and 80 female CLIL students were compared with 29 male and 39 female non-CLIL students, again from 10th grade. Regarding gender in CLIL and non-CLIL settings, statistically significant differences were found in the categories L2 self-confidence and English anxiety between CLIL females and CLIL males (females had higher anxiety and lower self-confidence), CLIL females and non-CLIL females (non-CLIL females had higher anxiety and lower self-confidence) and, in the case of English anxiety, non-CLIL females and non-CLIL males (non-CLIL females had higher anxiety). Thus, non-CLIL females had the lowest L2 self-confidence and highest anxiety, followed by the CLIL females, and then CLIL and non-CLIL males, with the CLIL males showing the highest self-confidence and lowest anxiety. Finally, Gallardo-del-Puerto and Blanco-Suárez (2021), as discussed in Section 3.6, compared the language learning motivation of male and female CLIL ($n = 124$) and non-CLIL ($n = 128$) primary level students by means of an adapted version of the AMTB. Findings indicated that while non-CLIL male students had significantly lower motivation than non-CLIL female students in overall motivation and intrinsic motivation, this was not the case in CLIL. The researchers thus suggest that, in terms of motivation, CLIL may provide a levelling effect and consequently a more egalitarian educational setting.

In conclusion, some studies have indeed found the suggested levelling effect of CLIL on gender, with statistically significant differences in non-CLIL students but no such difference in CLIL students. This was the case for both attitudes in Merisuo-Storm (2007) and motivation in Gallardo-del-Puerto and Blanco-Suárez (2021). Though no non-CLIL groups were assessed, Fernández Fontecha (2014a) also found no statistically

significant difference between male and female CLIL students in terms of motivation, while Fernández Fontecha (2014b) found no difference in terms of extrinsic motivation but statistically significant differences in terms of intrinsic motivation. Fernández Fontecha and Canga Alonso (2014), meanwhile, also found no statistically significant difference between male and female learners' motivation in CLIL, but also found the same in non-CLIL. On the other hand, other research has shown no evidence of a levelling effect, with female students outperforming male students in both CLIL and non-CLIL, for example, Lasagabaster (2008) for listening, writing and reading skills (higher results also for speaking, though not statistically significant); and Lasagabaster and Sierra (2009) for language attitudes. Sylvén and Thompson (2015) have, however, also found females to have higher scores in the more negative motivation categories such as anxiety and self-confidence. Their results found non-CLIL female students to report the highest English anxiety and lowest L2 self-confidence, followed by CLIL females, non-CLIL males, and finally CLIL males, who had the lowest anxiety and highest L2 self-confidence. These results indicate a clear difference both between genders and between students in different learning contexts. Other studies have shown differences between the male and female students in different aspects of motivation. For example, Heras and Lasagabaster (2015) revealed that while males had higher Ought-to Self than females in CLIL, females had higher Ideal L2 Self than males in non-CLIL. Thus, despite some findings suggesting the levelling effect of CLIL on gender, counterevidence shows that in many cases, females continue to outperform males in a CLIL context, both in terms of language skills, attitudes and motivation (both positive and negative aspects). As Lasagabaster and Sierra (2009) suggest, this may be attributed to the more limited form of CLIL at hand, as opposed to immersion which has shown a greater effect. However, as mentioned above, some studies have shown a degree of variation between male and female students in terms of the type or aspect of motivation (e.g., Fernández Fontecha, 2014b; Heras & Lasagabaster, 2015; Sylvén & Thompson, 2015). Given the supposition that a blurring of gender differences may arise due to male students' interest in the subject content (Heras & Lasagabaster, 2015), it would be beneficial for studies to analyse students' specific motivation and interest towards the content studied in CLIL in order to determine whether there is a consequential increase in language proficiency and language learning motivation.

Chapter 5: Content and Language Integrated Learning

This chapter will open with an overview of CLIL, outlining its background and key characteristics and then detailing in turn the main considerations in adopting the approach and some important issues regarding the language of instruction. The second section will provide a contextual background to CLIL as relevant to this study, outlining its implementation throughout Spain in general and also specifically within the autonomous community of La Rioja. The final section will provide an in-depth overview of the research that has been carried out on CLIL, focusing on the three key areas in this work: vocabulary, motivation and gender.

5.1. An Overview of Content and Language Integrated Learning

This section offers a theoretical overview of CLIL. It will first provide a description of what CLIL entails and explain how it differs from other similar approaches which combine content and language. It will then provide a discussion of the main advantages and disadvantages of adopting a CLIL approach.

5.1.1. What is Content and Language Integrated Learning?

The practice of teaching academic content through the medium of a language other than one's mother tongue is by no means a recent undertaking. Sylvén and Ohlander (2014), for example, point out that this approach goes as far back as the Roman Empire, where Latin, as the language of the ruling class, was used in educational settings throughout the empire. More recently, the period from 1984 to 1994 saw increasing interest in the various forms of bilingual practice, which aimed either to find coping strategies for those studying a language which was unfamiliar to them or to improve language learning itself (Marsh, 2009). Such various forms of combining content and language in the classroom have been addressed using a wide array of terms such as Immersion Education, Bilingual Teaching, Content Based Language Teaching (CBLT), Content-Based Instruction (CBI), English-Medium Instruction (EMI) and CLIL, to name but a few (Dalton-Puffer & Smit, 2007; Lightbown, 2014; Brown & Bradford, 2017). To this day, there remains a great deal of terminological confusion surrounding these various forms of teaching, with varying differences on each side of the Atlantic (Thompson & McKinley, 2018). The following two sections will clarify some of the similarities and differences between these different approaches, first highlighting geographical

preferences in terminology and outlining the European understanding of CLIL, and then clarifying some of the main differences between CLIL and EMI.

5.1.1.1. CLIL, CBI or Immersion. Within a North American context, the term CBI has been used as an umbrella term to describe any approaches which teach subject-matter in an L2 or FL (Brown & Bradford, 2017). According to Graham, Choi, Davoodi, Razmeh and Dixon (2018), CBI then comes in different forms, the most common of which are EMI and CLIL. In contrast, some European researchers, such as Dalton-Puffer, Llinares, Lorenzo and Nikula (2014), have claimed that it is impossible to classify CLIL as a type of CBI, given that CLIL is scheduled as content lessons, taught by content-trained teachers, and assessed as content. Cenoz (2015), however, argues that following an analysis of the characteristics CBI and CLIL, this proposition cannot be sustained. She points out that both CBI and CLIL programmes are timetabled as content classes and taught by content teachers and that, while different bodies may have different terminological preferences, CBI and CLIL essentially refer to the same thing. Others have also suggested that CLIL and CBI refer to the same concept but suggest that while the former is used in Europe, the latter is used in North America (Thompson & McKinley, 2018).

San Isidro (2019) has also highlighted the disparate views of CLIL conceptualisation, specifically regarding CLIL and previously existing immersion programmes. He notes how, on the one hand, Bruton (2013) criticises the approach as vague and too heterogeneous, claiming that there is no distinct difference between CLIL and previous immersion programmes. On the other hand, while CLIL is often linked to and identified with earlier bilingual education models (San Isidro, 2019) and shares characteristics with them (Somers & Surmount, 2011), it has been argued that the specific European context where CLIL takes places entails a fundamental difference. Gallardo del Puerto, Gómez Lacabex and García Lecumberri (2009) have similarly emphasised the difference in the sociolinguistic and sociocultural context of English-focused CLIL programmes in Europe as opposed to Canadian immersion or American bilingual programmes in terms of the authenticity of input: while in Europe the TL is generally not one that is used on a regular basis outside of class, in North America students often receive authentic extramural input. Similar reasoning is put forward by Nikula (2017), who notes that while immersion and CBI involve teaching students through their L2, in CLIL the language used is generally English (or another lingua franca), rather than an L2 which is

used in the learners' environment. CLIL may then be understood as "the European label for bilingual education" (Lorenzo, 2007, p. 28).

In Europe, the umbrella term CLIL was coined in the mid-1990s against the backdrop of European language policy promoting intercultural competence, multiple identities, and multilingual citizens (Dalton-Puffer et al., 2014). CLIL, AICLE (Aprendizaje Integrado de Contenidos y Lenguas Extranjeras) in Spanish, and EMILE (l'Enseignement de Matières par l'Intégration d'une Langue Étrangère) in French, are acronyms which have become firmly embedded in modern language teaching (Pérez Cañado & Ráez Padilla, 2015). It has been described as "any dual-focused educational context in which an additional language, thus not usually the first language of the learners involved, is used as a medium in the teaching and learning of non-language content" (Marsh, 2002, p. 15). Coyle's (2007, p. 545) definition also highlights the balance in this dual focus, referring to it as "an integrated approach where both language and content are conceptualised on a continuum without an implied preference for either". The approach is found from kindergarten to tertiary level and may range from the occasional use of FL texts in a given subject to covering the whole curriculum (Dalton-Puffer, 2007). Although CLIL implementation evidently varies from programme to programme, Dalton-Puffer et al. (2014) provide three prototypical characteristics of CLIL based on the empirical research to date:

1. CLIL languages are predominantly major or minor international *linguae francae* (in Europe English, French, Spanish, German), with a dominance of English.
2. CLIL happens alongside FL teaching, not instead of it.
3. CLIL is timetabled as content lessons and taught by content-trained teachers.

5.1.1.2. CLIL versus EMI. Before providing greater detail on CLIL in the next sections, it is also worth clarifying some differences between CLIL and EMI, which has also become increasingly popular in recent years both in Europe and on a global scale. EMI has been defined as "the use of the English language to teach academic subjects in countries or jurisdictions where the first language (L1) of the majority of the population is not English" (Dearden, 2014, p. 2). As is explained by Pérez-Vidal (2014, p. 36), its popularity has largely been motivated by "the increasingly competitive recruitment process of universities and the mobility policies within the European Union", as well as economic factors, with universities "trying to attract fee-paying students". While both CLIL and EMI involve teaching content through a TL, CLIL notably does not strictly

mention which particular language is to be used, while EMI evidently refers explicitly to the use of English. In addition, while the aim of CLIL is to further both content and language simultaneously, EMI does not necessarily have this objective (Dearden, 2014). Thus, CLIL typically involves learning subjects such as geography, physical education or music through any FL, with the dual focus of improving both content and language, whereas EMI is quite often used to describe tertiary level classes whereby English is used simply as a means of transferring content, without any particular focus on improving the language, for example, German-native university students studying a degree in economics through the medium of English.

5.1.2. Advantages and Disadvantages of CLIL

This section will outline the main considerations in adopting a CLIL approach. It will indicate the main benefits of CLIL and reasons why it has been increasingly implemented across educational institutions; outline the potential shortcomings of the approach; and focus on one potentially detrimental issue in the approach, namely the predominance of English and neglect of other languages in both practice and research.

5.1.2.1. Reasons for adopting a CLIL approach. Within the European context, the 1980s and 1990s saw a clear need for language teaching innovation, given the increasing need for L2 competencies in an internationalised world and the general dissatisfaction with ineffective L2 learning practices at the time (Goris et al., 2019). Against this backdrop, the enthusiasm for the implementation of CLIL largely came about as a result of the 1995 White Paper entitled *Teaching and Learning: towards the Learning Society* (Eurydice, 2006), in which the European Commission stated that “everyone should be able to communicate in two European languages in addition to their mother tongue” (Lasagabaster & López Beloqui, 2015, p. 42). In order to achieve this so-called 2 + 1 principle, the Resolution of the Council proposed, among other things, “the teaching of classes in a foreign language for disciplines other than languages, providing bilingual teaching” (Eurydice, 2006, p. 8). According to Dalton-Puffer and Smit (2007), there are three main arguments why the approach may be beneficial:

1. It provides conditions for more authentic language use.
2. It provides students with a purpose for language use in the classroom.
3. It increases exposure to the TL.

Firstly, as Dalton-Puffer (2007) highlights, CLIL classrooms may be a clever and economical way of providing a naturalistic environment, turning classrooms into streets and leaving behind the typical toils of the FL classroom. This is particularly so in cases where the TL is an FL as, being unable to find such learning opportunities outside of the classroom, the CLIL environment allows students to use the language in more authentic scenarios. Secondly, CLIL allows students to engage in real-life content-learning tasks with a specific purpose and has even been referred to as the “ultimate communicative methodology” (Graddol, 2006, p. 86). For example, Ioannou Georgiou (2012) highlights how students may use the TL in a geography class to create a map or in a science class to conduct an experiment. According to Larsson (2001), such task-based learning has the virtue of improving communicative skills, positively affecting social interaction, and encouraging students to gain a deeper sense of understanding. Finally, in addition to their regular language classes, students receive increased exposure to the TL via other content classes. As Lasagabaster and López Belouqui (2015) explain, they have the opportunity to use their FL in different contexts, allowing them to develop their intercultural and language skills while at the same time learning the curriculum content. In other words, there is a sense that students get “two for the price of one” (Zydatiβ, 2012, p. 27), increasing their exposure time to the language without placing extra demands on already busy timetables.

In addition to the above benefits, CLIL methodologies also have the advantage of uniting learning theories, language learning theories and intercultural understanding (Coyle, 2008). This multifaceted benefit is exemplified in Coyle’s (1999) 4Cs Conceptual Framework, discussed previously in Section 3.5.3, which states that CLIL has four key, interrelated parameters, namely Content (subject matter), Communication (language), Cognition (learning and thinking), and Culture (social awareness of self and “otherness”). Thus, CLIL methodologies ideally consider the relative value of these parameters for effective progression in each, and in doing so offer a teaching context in which language and content are integrated, where scaffolds are provided for both, and where cultural awareness and intercultural understanding are also added to the fold (Coyle et al., 2010). CLIL also has the potential advantage of providing students with a more diverse range of communicative skills, given its crucial influence on BICS (Basic Interpersonal Communicative Skills) and CALP (Cognitive Academic Language Proficiency) (Ball, Kelly & Clegg, 2016).

The BICS/CALP distinction was first introduced by Cummins (1979), who

defines BICS as conversational fluency and CALP as the ability to communicate concepts and ideas relevant to school success in both oral and written modes (Cummins, 2013). BICS and CALP have also been used interchangeably with the terms conversational and academic language proficiency (Khatib & Taie, 2016). According to Ball et al. (2016), the CLIL approach, and in particular content-led CLIL, is CALP-rich, and when carried out effectively can harness CALP and practise and balance it alongside the gentle influence of BICS. For example, as is indicated by Anderson (2011), by providing students with the opportunity to carry out communicative tasks such as writing a lab report in their TL, CLIL approaches offer learners the chance to develop their CALP, in addition to BICS, in a way that may otherwise not be possible.

In general, research into the CLIL approach has been quite positive, with most studies finding higher L2 levels for students enrolled in CLIL classes as compared with those in conventional FL classes (Goris et al., 2019). Dalton-Puffer (2008) offers some of the earliest insights into the specific language competencies which may be developed in CLIL (Table 5.1), which are discussed in turn below.

Table 5.1

Language Competencies Favourably Affected or Unaffected by CLIL

Favourably affected	Unaffected or Indefinite
Receptive skills	Syntax
Vocabulary	Writing
Morphology	Informal/non-technical language
Creativity, risk-taking, fluency, quantity	Pronunciation
Emotive/affective outcomes	Pragmatics

Note. From “Outcomes and processes in content and language integrated learning (CLIL): Current research from Europe” by C. Dalton-Puffer, in W. Delanoy & L. Volkman (Eds.), *Future perspectives for English language teaching* (p. 143), 2007, Carl Winter. Copyright 2008 by Carl Winter.

Regarding receptive skills, Prieto-Arranz, Fabra, Calafat-Ripoll and Catrain-González (2015, p. 125) remark that this advantage is to be expected, given that CLIL is intended to provide “comprehensible input + 1”, or “input which the learner can understand even

though it is still one step beyond their current ability in the target language”. It has also been highlighted that in CLIL settings receptive skills, and particularly reading, are worked much more actively than productive skills (Coonan, 2007). In addition, CLIL may understandably benefit comprehension skills by “enlarging the number of different speakers which learners are confronted with face-to-face” and potentially providing learners with “additional reasons for reading” (Dalton-Puffer, 2008, p. 6). Vocabulary, and in particular technical, semi-technical and general academic language, has also been found to be favourably affected, perhaps due to the fact that vocabulary is often the only linguistic skill which is treated explicitly in CLIL classrooms (Matiasek, 2005) (see [Section 5.3.1](#) for a comprehensive discussion of vocabulary research in CLIL). Regarding morphology, it has been suggested that aspects such as third person –s, irregular past tenses and models gain a higher degree of automatization and appropriacy of use, perhaps due to a parallel effect of time and quantity (Dalton-Puffer, 2008). As for speaking skills, Dalton-Puffer (2008) highlights that CLIL students exhibit greater creativity, risk-taking inclination, fluency and quantity. San Isidro (2019) suggests that these oral benefits may be attributed to the more active role that CLIL students play when using the language. This benefit is also suggested to go hand-in-hand with those regarding emotive/affective factors, given that learners may gradually lose their inhibitions to use the TL spontaneously in CLIL lessons (Dalton-Puffer, 2008). Motivation, in particular, has been discussed extensively in relation to CLIL, and will be dealt with in detail in Section 5.4.

San Isidro (2019, pp. 37-38) offers a more recent, comprehensive overview of the impact of CLIL on language learning. Regarding general proficiency, the overview cites and discusses over a decade of studies whose results show CLIL students outperforming non-CLIL students (e.g., Järvinen, 2005; Zydatiņ, 2007; Ackerl, 2007; Lasagabaster, 2008, 2009; Gallardo del Puerto et al. 2009; Ruiz de Zarobe 2010; Navés & Victori, 2010; San Isidro, 2010; San Isidro & Pérez Cañado & Lancaster, 2017; Lasagabaster, 2018; Pérez Cañado, 2018a). Regarding the specific skills, the overview highlights more mixed results. As in Dalton-Puffer (2008), oral skills, reading and vocabulary are seen to be positively affected. However, as San Isidro (2019) points out, studies investigating listening skills are somewhat less clear-cut, with some studies finding positive affects (Lasagabaster, 2008, 2011; Lorenzo, Casal & Moore, 2010; San Isidro, 2010) and others finding no significant differences (Navés, 2011). Writing, on the other hand, is suggested in fact to be favourably affected, particularly with regards to lexical and morphosyntactical resources, pragmatic awareness, and accuracy (San Isidro, 2019). For example, Ruiz de

Zarobe (2011) highlights results by Lasagabaster (2008), who found that in a bilingual context, CLIL students outperformed non-CLIL students in all language skills tested (reading, writing, speaking and listening), and most notably in writing and pronunciation, where gains were not expected. This author also notes findings from Navés and Victori (2010) and Navés (2011), which both find CLIL students to outperform non-CLIL students who are two to three years ahead of them: both studies found a CLIL advantage in terms of syntactic complexity, lexical complexity and fluency, while Navés (2011) additionally found advantages in terms in accuracy, lexical variety and error-free sentences. As can be seen, numerous studies have found advantages across a wide number of skills; however, there is still some degree of uncertainty regarding the benefits of CLIL in some areas, which will require further investigation. Both Ruiz de Zarobe (2011) and San Isidro (2019) have also highlighted the need for longitudinal studies in order to better understand the long-term effects of the approach.

In addition to the above benefits, CLIL may also be more advantageous to a wider range of students. As is explained by Dalton-Puffer (2008, p. 5), although students with special linguistics gifts can obtain a high proficiency via the average language classroom, “CLIL significantly enhances the language skills of the broad group of students whose foreign language talents or interest are average”. In other words, the approach may allow a higher number of language students the opportunity to improve their skills in a way which is not possible in the traditional language classroom. Finally, it has also been suggested that these benefits for the FL can be achieved without any detrimental effect on the students’ L1 and content learning (Merino & Lasagabaster, 2018a; Lasagabaster & Ruiz de Zarobe, 2010; Mehisto, Marsh & Frigols, 2008; Dalton-Puffer, 2008).

5.1.2.2. Potential shortcomings of a CLIL approach. Despite the many clear of advantages of CLIL, there are also evidently a number of issues to consider, with a number of researchers, such as Bruton (2011, 2012, 2013, 2015), Pérez Cañado (2011, 2012), Paran (2013) and Cenoz, Genesee and Gorter (2014) criticising the approach for being overly glamorised, highlighting its shortcomings, and questioning the validity of its outcomes (Goris et al., 2019). The main issues and criticisms are largely related to two areas: failings in CLIL research and the relationship between content and language.

5.1.2.2.1. Shortcomings in CLIL research. Regarding CLIL research, an initial issue is that CLIL has spread so rapidly that it has essentially outpaced measures of its impact (Pérez-Cañado, 2012). In other words, there has been such enthusiasm for the approach

that “practice has often preceded research” (Marsh, 2009, p. viii). This issue is also raised by Graddol (2006), who notes that CLIL has grown quite organically, rather than within top-down reform programmes. Baetens Beardsmore (2001, p. 10) summarises the practical consequences of this rapid increase in interest, which clearly comes alongside a number of obstacles:

The result has been that demand is outstripping supply for efficient, content-based multilingual education in many countries. Ever more parents are willing to plunge their children into education partly conducted in more than one language, but teacher availability, materials provision, curricular adjustment and exit criteria are struggling to keep up with rapid evolution.

In addition, although it has spread widely across Europe, in many cases CLIL has been unevenly implemented, with great variation in the legal frameworks in different countries and even regions within the same country (Frigols Martín, 2008). Thus, although CLIL programmes share many of the characteristics outlined by Dalton-Puffer et al. (2014) above, the term CLIL itself “encompasses more than a dozen educational approaches of bilingual education and CLIL programmes vary as much as European sociolinguistic and socioeducational contexts do” (Merino & Lasagabaster, 2018b, p. 20). This entails that interpretation of results is often only valid for the specific context at hand (Dalton-Puffer & Smit, 2013). Further issues pertain specifically to research design, including the scarcity of longitudinal studies and the lack of homogeneity of the groups under comparison (Pérez Cañado & Ráez Padilla, 2015). Firstly, a longitudinal perspective is necessary if we are to provide a meaningful interpretation of language learning, development, progress, change or gains (Ortega & Ibarra-Shea, 2005). However, according to Broca (2016), a significant and recurrent failing in CLIL research has been the lack of pre-tests at the outset. Pérez Cañado (2018a) and Goris et al. (2019) also point to the overwhelming lack of longitudinal CLIL research, which entails a shortage of conclusive evidence regarding the added value of CLIL in L2 learning. Secondly, Broca (2016) highlights that in many cases CLIL and non-CLIL students are different from the onset, indicating a lack of group homogeneity. In particular, differences have been highlighted with regards to language level, when there are prerequisites for taking CLIL or when more proficient students enrol in CLIL programmes, and motivation, given that admission to programmes is often voluntary and thus may inevitably attract more motivated students (Bruton, 2011). A case in point is the German context, where CLIL

students may receive two extra hours of English language classes for two years before starting, and an extra hour once they begin (Bruton, 2015). Such a scenario may unsurprisingly lead to a linguistic advantage for CLIL students even prior to starting the programme. For example, findings from Rumlich (2014) show the initial differences between pre-CLIL students, who had been taking extra language classes, and non-CLIL students, with the former predictably outperforming the latter. A similar situation arises in terms of differences in socioeconomic status (SES). Pérez Cañado (2018b), for example, has highlighted the important weight of SES on both L1 competence and content knowledge in CLIL settings, with results showing that students with a higher SES outperform those with a lower SES. These potential differences evidently call into question the validity of group comparisons when there is a failure to match the groups (Goris et al., 2019). Another issue with group homogeneity, highlighted by Jiménez Catalán and Agustín Llach (2017), is that CLIL groups often inevitably receive a higher number of hours of exposure to the TL than non-CLIL groups, given that CLIL classes are taken in addition to FL ones. In such cases, it is unclear whether positive effects are due to CLIL itself or to time. Due to these varying differences between groups, Bruton (2013) consequently argues that any observable benefits are attributable more to student selection and other factors than to the CLIL programme itself. Breidbach and Viebrock (2012) similarly suggest that the structural selectivity of CLIL may actually have a greater impact on student achievement than the type of instruction. It should be noted, however, that this selectivity may vary a great deal both in different countries and different regions, with CLIL in some places (e.g., Italy and some regions in Spain) being school-based rather than group-based (San Isidro, 2018). Thus, rather than a select number of students choosing to take CLIL, the whole school and all students must take part.

5.1.2.2.2. *Issues in the content and language balance.* Despite the presence of the word “integrated” in CLIL, the relationship between content and language has been characterised by tension and even conflict (Dalton-Puffer, 2007). One issue is the extent to which classes should focus on content and language. As is pointed out by Cenoz et al. (2014), the dual role of the two has often been understood in different ways, leading to an unclear idea of how much of focus should be placed on each one. Even within Europe, CLIL has been used as a broad term to describe a wide variety of forms in school practices, with some programmes adopting a hard, or content-led, CLIL and others adopting a soft, or language-led, CLIL (Ball et al., 2016). Ball (2009), for example, has provided a

understood and implemented CLIL, with the former viewing English as a hindrance, getting in the way of delivering content while the latter viewed the CLIL classes as an opportunity for students to improve their different skills in English. These findings further highlight the content/language dichotomy, whereby focus is placed on either one or the other rather than adopting an integrated approach (Karabassova, 2018), and emphasise the diversity of CLIL practices even within the same educational context.

A second issue is that integrating content and language evidently places additional demands on teachers. The push for adopting CLIL has not only outpaced research, as noted above, but also training for the teachers who are expected to deliver it (Deller & Price, 2007). As a result, content teachers may not be adequately proficient in the TL, while language teachers may be unfamiliar with the content. Additional challenges may also be found with developing methodologies, as introducing CLIL also entails “a methodological *revolution*”, navigating from subject methodologies and language teaching methodologies towards CLIL-specific methodologies (Pavón Vásquez & Rubio, 2010, p. 48). Tabuenca Cuencas and Alcaraz Mármol (2014) also caution the extra demands placed on students, as the change from acquiring content in their L1 to their L2 may inevitably increase difficulty in content acquisition, possibly leading to student frustration and lack of motivation.

A final shortcoming concerns the language that is used in CLIL programmes which, given its central importance in this project, will be discussed in detail in the next section.

5.1.2.2.3. CLIL or CEIL: the “empirical vacuum”. Although the original objective behind adopting a CLIL approach was to spread multilingualism, there has been a remarkable and undeniable predominance of the use of English throughout CLIL programmes (San Isidro, 2018). In fact, some definitions of CLIL, such as that provided by Graddol (2006, p. 86), completely disregard the notion of using languages other than English (LOTEs), stating explicitly that CLIL is an approach that uses English: “CLIL is an approach to bilingual education in which both curriculum content – such as Science or Geography – and *English* [emphasis added] are taught together”. Lyster and Ballinger (2011) offer a similar “English-only” profile of CLIL when distinguishing between immersion and CLIL, indicating that while both aim to integrate content and language instruction, immersion targets languages other than only English. This issue has been highlighted by Dalton-Puffer (2011) who argues that, outside English-speaking countries, the prevalence of English is so overwhelming that it may make more sense to speak of

CEIL: Content and English Integrated Learning.

This dominance evidently also crosses over into research, as given that the majority of programmes are conducted in English, it stands to reason that the bulk of research to date has focused on CEIL. Cenoz et al. (2014, p. 257), for example, highlight how “much, if not most, research on CLIL has been conducted by ESL/EFL scholars”, while Pérez, Lorenzo and Pavón (2016, p. 485) highlight the “empirical vacuum” in the field, drawing attention to the absence of research into CLIL in LOTEs. In fact, as early as a decade ago, Dalton-Puffer et al. (2010, p. 288) called for comparative research across additional languages, in order to provide insight into the strengths and weaknesses of CLIL language-independently. Similarly, Cenoz et al. (2014) have called for a more critical, empirical examination of CLIL in diverse contexts, focusing not just on ESL/EFL but also on other L2s which have been largely neglected.

In order to provide such comparative research, there are three possible avenues that can be taken:

1. Compare programmes from different educational institutions which have different TLs, for example, one school which has CLIL in English and a different one which has CLIL in French.
2. Compare programmes from the same educational institution in which different students have the option to take CLIL in different TLs, for example, a school where students have the choice to take CLIL in either English or French.
3. Compare trilingual CLIL programmes in which students are enrolled in CLIL classes in two L2s, for example, a school where all students take CLIL in both English and French.

The first two options compare different bilingual programmes which carry out CLIL in the state language and one other language. This may involve comparing programmes from different educational institutions which have different TLs; evidently the most problematic option, given that both different programmes and groups are being compared. Alternatively, comparisons could be made within programmes in the same institution when there is a choice to take CLIL in one of two different L2s. An example of this can be seen in van Mensel, Hiligsmann, Mettwie and Galand (2020), where English and Dutch CLIL students in French-speaking Belgium are compared. While preferable to the previous scenario, this option is still somewhat problematic, given that different groups of students are being compared. A potentially more fruitful approach is comparing

students enrolled in trilingual CLIL programmes, so as to be able to compare the same participants enrolled in the same programme, but in two different languages. For example, Baten et al. (2020) recently carried out the first such comparison in Flemish learners of English and French as L2s (see [Section 5.3.2](#) for an in-depth discussion).

With respect to European trilingual CLIL programmes, there are two main trilingual CLIL scenarios available: those which combine the national language and two FLs, and, more commonly, those which combine the national language, a minority language and an FL (Eurydice, 2006). In the 2004/2005 academic year, Eurydice (2006) indicated seven European countries which provided scope for trilingual CLIL: Austria, Estonia, Latvia, Luxembourg, the Netherlands, Spain and Sweden. Out of these, only Spain and Latvia provided scope for combining the national language and two FLs, with the other five combining the national language, an FL and a minority language. Data collected ten years later (Eurydice, 2017) shows a huge increase in the potential for trilingual CLIL, with 17 countries providing content classes in the state language and at least two other types of language (Table 5.2). It should be noted that although a country may provide CLIL in two other types of language, it does not necessarily imply that programmes include both languages simultaneously. In addition, the data provided by Eurydice does not include scenarios in which there are two co-official languages and one or more FLs (for example, Spanish, Catalan, English and French) or in which there is one state language and two FLs (for example, Spanish, English and French), the latter of which is the case in the current project carried out in Spain.

However, despite the existence of such programmes, there is still a clear lack of studies dealing with the effects of three languages in CLIL (Merino & Lasagabaster, 2018a), leading researchers to question both the efficacy of trilingual programmes and, in particular, the effect that the increasing presence of English has on the development of minority or less spoken languages (Lasagabaster & Sierra, 2009).

Table 5.2*Types of CLIL Provision in Europe*

Number of languages	Type(s) of language	Countries
State language + at least two other types of language	Foreign language + Regional, minority or territorial language	Czech Republic, Germany, Estonia, Spain, France, Italy, Latvia, Lithuania, Hungary, Austria, Poland, Slovakia, Sweden, former Yugoslav Republic of Macedonia
	Foreign language + Another state language	Belgium (French and Flemish communities), Switzerland
	Foreign language + Another state language + Non-territorial language	Finland
State language + one type of language	Foreign language	Bulgaria, Denmark, Cyprus, the Netherlands, Portugal, Romania, England, Liechtenstein, Norway, Serbia
	Regional, minority or territorial language	Croatia, Slovenia, Wales, Northern Ireland, Scotland
	Another state language	Belgium (German community), Ireland, Luxembourg, Malta

Note. Adapted from *Key Data on Teaching Languages at School in Europe – 2017 Edition* (pp. 55-56), by Eurydice, 2017, Publications Office of the European Union. Copyright 2017 by Education, Audiovisual and Culture Executive Agency.

5.2. The Context

The following two sections provide an overview of the CLIL context, firstly throughout Spain in general and then in the specific autonomous community where the study is carried out, La Rioja.

5.2.1. CLIL in Spain

In order to understand the truly rich diversity of Spanish CLIL, it is necessary first to outline its geographic, linguistic and educational make-up. In Spain, there are 17 autonomous regions as well as the autonomous cities of Ceuta and Melilla, with education regulated nationally by the 2013 *Ley Orgánica para la Mejora de la Calidad Educativa* (Organic Law for the Improvement of the Quality of Education) (Manzano Vásquez, 2015) and more recently its replacement, the 2020 *Ley Orgánica por la que se modifica la LOE de 2006* (Organic Law to Modify the 2006 Organic Law of Education) (BOE, 2020). Importantly, unlike its predecessor, this updated law does not stipulate that Spanish must be the language of instruction, thus giving students the right to receive instruction in Spanish or in any other co-official languages in their respective autonomous communities, provided that they finish obligatory education with a full command of Spanish and of the co-official language in their region (BOE, 2020). Though this law regulates education on a national level, the 1978 democratic constitution granted the autonomous communities political and administrative power (Frigols Martín, 2008), thus allowing them to adapt this law to their own territories (Ruiz de Zarobe & Lasabagaster, 2010). This decentralisation of Spanish education entails a great deal of variation in each community, with each one adopting different models depending on the particularities of the area (Fernández Fontecha, 2009). As is highlighted by Ruiz de Zarobe and Lasabagaster (2010), these different CLIL models can be classified into two main contexts: monolingual communities and bilingual communities. While in monolingual communities education is carried out in the official language, Spanish, and one or two FLs, bilingual communities use Spanish in conjunction with another co-official regional language (Basque, Galician, Catalan or Valencian) and/or one or two FLs (Caraker, 2016). Eurydice (2017) provides specific information regarding the status of these different languages, as summarised in Table 5.3. According to Frigols Martín (2008), these different modes allow for a variety of scenario-dependant benefits, including promoting bilingualism in a monolingual community, fostering multilingualism in an already bilingual community, and improving competence in English.

Table 5.3*CLIL Instruction through Different Languages in Spain*

Language status	Languages	ISCED Level
1 state language + 1 or 2 foreign languages*	Spanish + English; French; German; Italian; Portuguese	1-3
1 state language + 1 regional/minority language with official language status	Spanish + Basque; Catalan; Galician; Occitan; Valencian	1-3
1 state language + 1 regional/minority language with official language status + 1 foreign language	Spanish + Basque + English; French Spanish + Catalan + English; French Spanish + Galician + English; French; Italian; Portuguese Spanish + Valencian + English; French; Italian; Portuguese	1-3
	Spanish + Catalan + Portuguese	2-3

Note. Adapted from *Key Data on Teaching Languages at School in Europe – 2017 Edition* (p. 161), by Eurydice, 2017, Publications Office of the European Union. Copyright 2017 by Education, Audiovisual and Culture Executive Agency.

*Though the data provided by Eurydice includes only one foreign language in this scenario, some CLIL programmes in Spain include two foreign languages, as discussed below.

As in other European countries, CLIL has been widely implemented throughout Spain, largely as a response to the European policies promoting multilingualism and an increased awareness of the need for FL learning (Ruiz de Zarobe & Lasagabaster, 2010). Goris et al. (2019, p. 695) have in fact suggested that in Spain “the CLIL approach was planted in fertile soil”. This is, firstly, due to the desperate need for an improved approach to language teaching given that, in the European Community, Spain has been found to be below the average in FL ability (European Commission, 2012). As pointed out by Manzano Vásquez (2015, p. 138), this negative finding has prompted Spanish education authorities “to reaffirm their commitment with multilingualism and take decisive steps to foster foreign language competences among students”. In light of this, CLIL has emerged as a promising solution to tackle this deficit. A second reason why the CLIL approach is well-suited to a Spanish context is the ample experience in dual-language content teaching in Spain’s bilingual regions. Though interest in CLIL has become especially popular in Spain since the turn of the century (Ruiz de Zarobe & Lasagabaster, 2010), in certain regions, such as the Basque Country (Sierra, 1994), Galicia (Vez, 2011) or Catalonia (Artigal, 1993), bilingual programmes are by no means a recent development (Manzano Vásquez, 2015). With the aim of normalising their official languages and ensuring additive bilingualism, some of these regions have implemented programmes since the 1980s (Pérez Vidal, 2007), and thus may provide a rich resource of experience from which to draw. This background has been beneficial both for bilingual and monolingual communities in adopting a CLIL approach with FLs, with the former having the possibility of transferring methodological procedures from regional to FLs, and the latter being able to draw on the experience of the bilingual communities (Ruiz de Zarobe & Lasagabaster, 2010).

Due to this rich linguistic diversity, Spain has quickly become one of the European leaders in CLIL practice and research (Coyle, 2010). In practical terms, numerous programmes have been adopted throughout the country and, although they have not been fully implemented in all autonomous communities, the vast majority of regions have carried out pilot and/or experimental programmes (Frigols Martín, 2008). Table 5.4 below, which draws from Ruiz de Zarobe and Lasagabaster (2010), Ortega-Martín and Trujillo (2018) and in particular Fernández Fontecha (2009), provides an overview of the main CLIL programmes in Spain and the autonomous communities where they have been implemented.

Table 5.4*CLIL Programmes in Spain*

Programme	Autonomous Communities
Bilingual Education Project (BEP) (The Spanish Ministry of Education and the British Council Project)	Aragón, Asturias, the Balearic Islands*, Cantabria, Castilla y León, Castilla La Mancha, Ceuta, Extremadura, Madrid, Melilla, Murcia, Navarra*
PALE (Programa de Apoyo a La Enseñanza y el Aprendizaje de Lenguas Extranjeras) (Foreign Language Learning and Teaching Support Programme)	Andalucía, Aragón, Asturias, the Canary Islands, Castilla y León, Castilla La Mancha, Catalonia*, Extremadura, Galicia*, Madrid, Murcia, Navarra*, La Rioja, Valencia*
Secciones Bilingües/Secciones Europeas (Bilingual Sections /European Sections)	Andalucía, Aragón, Asturias, the Balearic Islands*, the Canary Islands, Cantabria, Castilla y León, Castilla La Mancha, Catalonia*, Extremadura, Navarra, Galicia*, Madrid, Murcia, La Rioja, Valencia*
PILC (Proyectos de Innovación Lingüística en Centros) (School Language Innovation Projects)	La Rioja
ETC (English through Content)	Navarra*
Proyecto Bilingüe (Bilingual Project)	Madrid
Plan de Fomento del Plurilingüismo (Plurilingualism Promotion Plan)	Andalucía
Programa de educación bilingüe enriquecido (PEBE) (Bilingual Teaching Enriched Programme)	Valencia*

Note. * = bilingual communities or sectors

As can be seen, programmes can be divided into those larger ones which are implemented throughout numerous autonomous communities, and regional ones which are unique to just one autonomous community. In the former, the BEP, implemented in the academic year 1996-1997, is a joint bilingual project which aims to provide bilingual and bicultural education by combining both Spanish and English curricula (Fernández Fontecha, 2009). Taking the form of early and partial immersion, the BEP allocates around 40% of time to English and 60% to Spanish. Next, the PALE programme aims to provide linguistic support to teachers involved in CLIL teaching, by providing aid such as 200 hours of training in foreign-language teaching and offering a two-week SA visit (Fernández Fontecha, 2009). A final programme which has been widely implemented is the Bilingual Sections which, given that it is one of the main programmes being implemented in La Rioja, will be discussed in the following section. The latter group of programmes are those which are implemented only in one autonomous community. While these programmes have not been implemented in the same way in all autonomous regions, all have been accompanied by teacher training schemes which aim to assist teachers with the required linguistic and methodological skills to implement CLIL, such as language and methodology courses, and periods of SA (Ruiz de Zarobe & Lasagabaster, 2010).

In addition to these practical incentives, Spain has also been a forerunner in CLIL research, producing a wide body of research on the topic. Particularly noteworthy has been the 2010 volume *CLIL in Spain: Implementation, Results and Teacher Training*, edited by David Lasagabaster and Yolanda Ruiz de Zarobe, which offers a comprehensive overview of CLIL throughout Spain, dealing with the implementation and results of CLIL programmes, as well as teaching training, and offering advice on the necessary steps to be taken moving forward. Other clear contributions can be found in the 2010 volume *Content and Language Integrated Learning: Evidence from Research in Europe*, edited by Yolanda Ruiz de Zarobe and Rosa María Jiménez Catalán, which collates research findings from several different studies, the vast majority carried out by Spanish researchers in Spanish institutions. More recently, José Luis Ortega-Martín, Stephen Pearse Hughes and Daniel Madrid edited the 2018 volume *Influencia de la Política Educativa de Centro en la Enseñanza Bilingüe en España* [Influence of School Educational Policy on Bilingual Education in Spain], which included studies carried out across eight different autonomous communities and evaluated the organisation and quality of the CLIL approach in each region. Finally, Keiko Tsuchiya and María Dolores Pérez Murillo offer the 2019 volume *Content and Language Integrated Learning in*

Spanish and Japanese Contexts, focusing on policy, practice and pedagogy across a range of educational levels.

Such initiatives, both in practice and research, have been hugely important in the field, and have presented Spain as “a model for the multiple possibilities offered by the broader CLIL spectrum and thus for other countries seeking to implement it” (Pérez-Cañado, 2012, p. 327). The following section will outline the characteristics of CLIL practice and research in the specific Spanish region where this doctoral research is carried out: La Rioja.

5.2.2. CLIL in La Rioja

La Rioja is a small, monolingual autonomous community in the north of Spain, with a population of around 319,796 people (Instituto Nacional de Estadística, 2021). As in the rest of Spain, CLIL has become increasingly popular in recent years. In the academic year 2014-15, over 20,000 students, 30% of the total, were enrolled in some kind of bilingual programme (Manzano Vásquez, 2015). This number has steadily been increasing, with new schools joining these programmes each year (Gobierno de la Rioja, n.d.). Manzano Vásquez (2015) discusses three key bilingual initiatives that have been taken in La Rioja: the *Proyectos de Innovación Lingüística en Centros*, or PILC (School Language Innovation Projects), the *Programa de Competencia Lingüística Integrada en Centros* (School Integrated Language Competence Programme), and the Bilingual Sections. The following three sections outline the main characteristics of these initiatives.

5.2.2.1. School Language Innovation Projects. The PILC bilingual initiative was implemented in the academic year 2004-2005, and is a project carried out only in La Rioja (Consejería de Educación al Consejo Escolar, 2010). Upon its inception, the programme was available in three distinct modalities (Manzano Vásquez, 2015), as summarised in Table 5.5. This ranged from very limited use of the FL in Model A to studying an entire subject in Model C. This was later merged into one modality, PILC A, whereby three units or parts of the curriculum were studied in the TL, one in each semester, as in the previous Model B. Table 5.6. below provides an overview of the PILC programme’s implementation in the 2019/2020 and 2020/21 academic years, in order to demonstrate the trends and variances in recent years.

Table 5.5*School Language Innovation Projects Modalities*

	Use of foreign language
Model A	Routines, greetings, instructions, etc.
Model B	The contents of at least one unit or part of the curriculum in each semester.
Model C	The contents of one complete subject in the curriculum.

Table 5.6*Summary of the School Language Innovation Project in the 2019/2020 and 2020/2021 Academic Years*

2019/2020				
Language	Model B	Model C	Model B/C	Total
English	128	106	2	236
French	3	-	-	3
German	3	-	-	3
Italian	1	-	-	1
Total	135	106	2	243
2020/2021				
Language	PILC A			
English	205			
French	2			
Italian	2			
Total	209			

As shown, in the 2019/2020, 243 teachers in 60 different centres took part in the programme using either a Model B or Model C approach, across a range of educational levels. As can be seen, programmes were carried out overwhelmingly in English, accounting for 97% of the total. Furthermore, in no other language was Model C carried out. It should be noted, however, that in previous years, a higher number of teachers were using French in the programme, albeit via Model A. In addition, up until the 2017/2018 academic year, one centre in particular was carrying out French via Model C. In subsequent years, the centre was involved instead in the Bilingual Sections, discussed

below. More recently, in 2020/2021, a total of 209 coordinators and participants across 42 different educational centres took part in the new, single-modality PILC A. As in the previous year, the language chosen for the programme was almost always English (98%), though some participants opted for French (1%) or Italian (1%).

5.2.2.2. School Integrated Language Competence Programme. Another bilingual programme carried out only in La Rioja was the School Integrated Language Competence Programme (Gobierno de La Rioja, 2013). The programme aimed to develop the communicative competence of students and professional teachers. Teachers involved in the programmes, required to have a B1 of the CEFR, thus designed activities and tasks to be carried out in the FL in other content areas, with the aim of developing the students' competence in linguistic communication (Gobierno de La Rioja, 2013). The programme, however, was unfortunately short-lasting, starting in 2012-2013 with 11 schools, reducing to 4 schools in 2013-2014, and subsequently not being implemented (Manzano Vásquez, 2015).

5.2.2.3. Bilingual Sections. As indicated in Section 5.2.1, the Bilingual Sections, or European Sections, is an initiative which been rolled out on a national level, both in monolingual and bilingual communities, and in both primary and secondary schools (Fernández Fontecha, 2009). The programme was first implemented in 2008-2009 (Manzano Vásquez, 2015) and has become increasingly widespread throughout the region, with 26 schools involved in the academic year 2020/2021, 13 of which were public schools (5 primary and 8 secondary) and 13 of which of which were private (8 primary and 5 secondary) (*Consejería de Educación, Cultura, Deporte y Juventud*, 2021). The local government set out strict criteria for involvement in the project, including the following regarding implementation and organisation of the programme in secondary and vocational education (BOR, 2018, p. 7896):

Implementation

1. The project should begin in the first cycle of obligatory secondary education and continue throughout subsequent cycles and into upper secondary education or vocational education.
2. At least 30% of non-language subjects in the curriculum are carried out in the foreign language, including at least one core subject (bilingual projects) or two core subjects (plurilingual projects).

3. Centres involved gradually adapt their teaching until they have become bilingual or plurilingual at all levels.

Organisation

1. The language is used as an instrument for the acquisition of knowledge in other subjects. Students receive instruction in English every day, and, in the case of plurilingual projects, in the other foreign language on at least three days.
2. The English Department is responsible for coordinating the project. In the case of plurilingual projects, a teacher of the second foreign language, or the second foreign language department if the centre so decides, assists in the coordination.
3. Language classes are adapted to meet the specific curriculum of the bilingual project.
4. All areas in which the centre decides to teach in the foreign language are entirely in that language.

In addition, teachers are required to show that they have at least a B2 language level according to the CEFR as well as basic competencies in CLIL (BOR, 2018). As in the PILC programme, teachers are provided with a number of resources in order to help them achieve this, including having native language assistants and priority in access to languages courses (in particular those geared towards reaching and maintaining a C1 or C2 level of the CEFR) and methodological courses.

Having provided a theoretical and contextual background to CLIL in the first two sections, the following final three sections will now provide an overview of the most pertinent research that has been carried out within a CLIL context, focusing specifically on the three areas which are of central interest to this project: Vocabulary, Motivation and Gender.

5.3. Research on Vocabulary and CLIL

As noted above, vocabulary has been found to be positively affected in CLIL, particularly with regards to technical, semi-technical and general academic language. Baten et al. (2020) suggest a number of factors which may contribute to this positive effect of a CLIL approach on vocabulary development:

1. The interaction of explicit and implicit learning conditions
2. Meaningful context and more opportunities to practise
3. Non-threatening atmosphere, in terms of error correction
4. Potential for more student-teacher interaction
5. Reading and extra-curricular contact

Firstly, as is indicated by Merikivi and Pietilä (2014), CLIL is optimal for vocabulary learning, given that it combines both explicit and implicit learning conditions. This may occur through various activities such as reading different texts, listening to the teacher's instructions and explanations, and communicating with the teacher or other classmates (Tabuenca Cuevas & Alcaraz Mármol, 2014). Secondly, as is highlighted by Ruiz de Zarobe (2011), a CLIL approach clearly offers more opportunities to learn vocabulary, whether explicitly or implicitly, in a meaningful way. This is because by learning the target content, students receive a more contextualised opportunity to practise the language, given that it is used to fulfil real purposes (Rodríguez Bonces, 2012). This meaningful context is suggested to be extremely beneficial for vocabulary retention. As Tabuenca Cuevas and Alcaraz Mármol (2014) point out, this is closely related to Laufer and Hulstijn's (2001) Involvement Load Hypothesis, which states that the more involvement there is during word processing, the better the words will be retained. This hypothesis has been confirmed by a number of studies including those by Keating (2008), which found that beginners were better able to retain vocabulary when writing sentences as compared to only filling in blanks, and Lu (2013), which found blank-fill exercises to be more beneficial than reading alone. Thus, according to Tabuenca Cuevas and Alcaraz Mármol (2014), if students receive more meaningful learning opportunities and are thus more involved in the tasks at hand, they may be better able to retain the target vocabulary. Thirdly, it has been suggested that a CLIL classroom fosters a non-threatening atmosphere as, given that language learning is not the sole focus, students may feel less afraid to use the language and less concerned about making mistakes (Baten et al., 2020). This is because, rather than having linguistic accuracy be the central concern, the main priority is meaningful communication, whereby the learner focuses on the transmission of some kind of message or information. Fourthly, some researchers, such as Nikula (2010), report a higher amount of student-teacher interaction, which may also potentially mean a higher frequency of lexical encounters. Baten et al. (2020), however, note that a CLIL approach does not always guarantee that there will be more interaction (see, for

example, Lo and Macaro, 2015, who report little interaction in a CLIL approach). Thus, as in any approach, the actual amount of interaction is subject to variation and may differ from classroom to classroom. Finally, CLIL students may also come into contact with the language to a larger degree than non-CLIL students, as has been suggested and found in studies by Sylvén (2004, 2006, 2019). However, it should be noted that this author found that the CLIL students outperformed non-CLIL students in a receptive vocabulary test, even before the CLIL instruction has started. Thus, it is not necessarily the CLIL approach that led to vocabulary gains, but rather the students who enrolled in CLIL classes also reported a higher degree of extramural language contact and achieved a higher score on the vocabulary test.

Given these various factors, conducive to a favourable environment for vocabulary development, numerous researchers have set out to investigate the benefits of a CLIL learning environment on vocabulary. The following two sections offer an overview of these studies, first detailing those which focus on receptive vocabulary and then those which deal with productive vocabulary. Both sections focus specifically on CLIL in a Spanish context, however, other relevant research is also discussed. This research is of primary importance for this thesis, given its aim to analyse vocabulary, and in particular productive vocabulary, within a CLIL context.

5.3.1. Receptive Vocabulary in CLIL

Within a Spanish context, receptive vocabulary in CLIL has by and large been assessed using Schmitt et al.'s (2001) receptive VLT, generally at the 2000-word level. As shown in Table 5.7 and discussed below, these studies have been carried out with a variety of age groups, ranging from 4th grade to 10th grade. The majority of the studies deal with students whose L1 is Spanish and L2 is English, though some studies carried out in bilingual regions include additional languages, specified below whenever relevant.

Table 5.7*Summary of Research Investigating Receptive Vocabulary in Spanish CLIL*

Authors	Age group	Tests	Main Findings
Jiménez Catalán & Ruiz de Zarobe (2009)	Young learners (6 th grade) 65 CLIL + 65 non-CLIL	1K WT 2K VLT	CLIL students did significantly better on both receptive tests than the non-CLIL group.
Canga Alonso (2013a)	Young learners (6 th grade) 79 CLIL	2K VLT	Students' receptive vocabulary knowledge was below the 1000-frequency band.
Canga Alonso (2013b)	Young learners (5 th grade) 61 CLIL Teenagers (8 th grade) 194 non-CLIL	2K VLT	With the same time of exposure, the older, non-CLIL groups surpassed the younger CLIL groups in terms of receptive vocabulary size (statistically significant difference).
Fernández Fontecha (2014c)	Young learners (5 th grade) 55 CLIL Teenagers (8 th grade) 183 non-CLIL	2K VLT	With the same time of exposure, the older, non-CLIL groups surpassed the younger CLIL groups in terms of receptive vocabulary size (statistically significant difference).
Fernández Fontecha (2014a)	Young learners (5 th grade) 55 CLIL	2K VLT	Positive correlation between male students' intrinsic motivation and receptive vocabulary.
Fernández Fontecha (2014b)	Young learners (6 th grade) 66 CLIL	2K VLT	No statistically significant difference in the receptive vocabulary size of the male and female CLIL students; no correlation between receptive vocabulary and intrinsic and extrinsic motivation.

Table 5.7 (continued)

Fernández-Fontecha (2015)	Young learners (4 th grade) 58 CLIL Teenagers (7 th grade) 304 non-CLIL	2K VLT	With the same time of exposure, the older, non-CLIL groups surpassed the younger CLIL groups in terms of receptive vocabulary size (statistically significant difference).
Canga Alonso (2015a)	Young learners (5 th grade) 61 CLIL + 194 non-CLIL	2K VLT	All students' receptive vocabulary was lower than 1000 words; statistically significant differences found in favour of CLIL learners.
Canga Alonso (2015b)	Young learners (6 th grade) 79 CLIL + 199 non-CLIL Teenagers (10 th grade) 132 non-CLIL	2K VLT	CLIL group had a statistically significant higher receptive vocabulary than the non-CLIL group; with the same time of exposure, the older, non-CLIL groups surpassed the younger CLIL groups in terms of receptive vocabulary size (no statistically significant different).
Agustín Llach & Canga Alonso (2016)	Young learners (4 th to 6 th grade) 58 CLIL + 49 non-CLIL	2K VLT	Learners' receptive vocabulary was within the most frequent 1000 words; CLIL learners outperformed their non-CLIL peers (statistically significant difference in 5 th and 6 th grade).
Arribas (2016)	Teenagers (10 th grade) 73 CLIL + 19 non-CLIL	2KVLT 3K VLT	The CLIL group had a higher score in the receptive vocabulary tests, though the differences were not statistically significant.

Table 5.7 (continued)

Iglesias Diéguez & Martínez- Adrián (2017)	Teenagers (7 th grade) 15 CLIL + 10 non- CLIL Teenagers (9 th grade) 15 CLIL + 15 non- CLIL	1K VLT 2K VLT	CLIL students outperformed non- CLIL students in the same year; younger CLIL students also performed equally as well as the older non-CLIL learners who had received approximately the same number of hours of exposure.
Castellano- Risco (2018)	Teenagers (9 th grade) 24 CLIL + 20 non- CLIL	Yes/No test Vocabulary learning strategies test	CLIL students outperformed non- CLIL students; clear differences between the groups in the vocabulary strategies test: CLIL learners made significantly more use of consolidation strategies and non-CLIL learners used more discovery strategies.
Castellano- Risco et al. (2020)	Teenagers (9 th grade) CLIL + non-CLIL	2K VLT	Statistically significant difference both at the 2K band and the academic vocabulary band in favour of CLIL learners; three CLIL subgroups did not present statistically significant differences in their receptive vocabulary, despite having different lengths of exposure.

With regards to primary levels, Canga Alonso (2015a) compared 61 CLIL and 194 non-CLIL students in 5th grade using the 2000-frequency band of the receptive VLT. Results found all students' receptive vocabulary to be lower than 1000 words; however, there were statistically significant differences in favour of CLIL learners. As the author notes, this advantage may be attributed to the longer exposure to English received by the CLIL learners. Fernández Fontecha (2014a) analysed the effect of gender and motivation on

receptive vocabulary size in 55 5th grade CLIL students. Results revealed a positive correlation between male students' intrinsic motivation and their receptive vocabulary, finding that male students who were more intrinsically motivated also had a higher receptive vocabulary. Fernández Fontecha (2014b), Canga Alonso (2013a, 2015b), Agustín Llach and Canga Alonso (2016) and Jiménez Catalán and Ruiz de Zarobe (2009) all dealt with students in 6th grade of primary education. As in Fernández Fontecha (2014a), Fernández Fontecha (2014b) investigated the effect of gender on receptive vocabulary size and motivation, this time with 66 CLIL learners in 6th grade. Results showed no statistically significant difference in the receptive vocabulary size of the male and female students. Canga Alonso (2013a) analysed the receptive vocabulary of 79 CLIL students using the 2000-frequency band of the receptive VLT. As in the case of the 5th grade students discussed above, results showed that the students' receptive vocabulary knowledge was below the 1000-frequency band, which the author suggests may entail difficulty in understanding spoken and written discourse in English. Canga Alonso (2015b) compared these same 79 CLIL students with 199 6th grade non-CLIL students, and reports that the CLIL group had a statistically significant higher receptive vocabulary than the non-CLIL group. A subgroup of the above learners was also assessed by Agustín Llach and Canga Alonso (2016), who compared CLIL ($n = 58$) and non-CLIL ($n = 49$) students using the 2000-frequency band of the receptive VLT. Using a longitudinal approach, the authors tested the participants across three consecutive years (4th, 5th and 6th grades). Results found learners' receptive vocabulary to be within the most frequent 1000 words and indicated that CLIL learners outperformed their non-CLIL peers, a result which was statistically significant in 5th and 6th grade. Finally, Jiménez Catalán and Ruiz de Zarobe (2009) analysed 130 female students (65 CLIL with Spanish L1, Basque L2 and English L3 and 65 non-CLIL with Spanish L1 and English L2), using a language-level cloze test, the 1000-word receptive test and the 2000-frequency band of the receptive VLT. Results showed that the CLIL group did significantly better on both receptive tests than the non-CLIL group. However, as in Canga Alonso (2015a), the authors acknowledge that, given that the CLIL group received a higher number of hours of instruction, it is not possible to determine whether these results are due to CLIL itself or time of exposure.

Following from this issue regarding the consequential higher number of hours of exposure on the part of CLIL groups, a number of researchers have compared young CLIL learners with older non-CLIL learners who had received an equal number of hours of

exposure, all using the 2000-frequency band of the receptive VLT. Fernández Fontecha (2015) compared 4th grade CLIL students ($n = 58$) with 7th grade non-CLIL students ($n = 304$), all of whom had received 734 hours of English; Canga Alonso (2013b) compared sub-groups of the same participants the following year when the students were in 5th grade ($n = 61$) and 8th grade ($n = 194$); Fernández Fontecha (2014c) compared 5th grade CLIL students ($n = 55$) and 8th grade ($n = 183$) non-CLIL students, all of whom had received 839 hours of English; and Canga Alonso (2015b) compared the 79 6th grade CLIL students and 199 6th grade non-CLIL students discussed above with 132 10th grade non-CLIL students. In all studies, results showed that with the same time of exposure, the older, non-CLIL groups surpassed the younger CLIL groups in terms of receptive vocabulary size. These results were statistically significant in all studies except Canga Alonso (2015b), who reported only a slight, non-significant difference between these two groups of students.

The remaining studies summarised above focus solely on teenagers. Iglesias Diéguez and Martínez-Adrián's (2017) sample consisted of 55 Basque-Spanish learners of English, who were divided into four groups: 7th grade CLIL ($n = 15$) and non-CLIL ($n = 10$), and 8th grade CLIL ($n = 15$) and non-CLIL ($n = 15$). Participants were assessed using the Quick Placement Test and the 1000- and 2000-frequency bands of the receptive VLT. The results showed that CLIL students outstripped non-CLIL students in the same year, but also that the younger CLIL students performed equally as well as the older non-CLIL learners who had received approximately the same number of hours of exposure. Castellano-Risco (2018) compared 24 CLIL and 20 non-CLIL students in 9th grade, using the Yes/No test and a vocabulary learning strategies test. The results show a CLIL advantage, however, the author highlights that this may have been due to the higher amount of exposure received by the CLIL students, the fact that learners had to pass an exam beforehand, or the use of CLIL instruction. Results of the vocabulary strategies test also show clear differences between the groups, with CLIL learners making significantly more use of consolidation strategies and non-CLIL learners using more discovery strategies. Castellano-Risco, Alejo-González and Piquer-Píriz (2020) also assess 9th grade students, comparing 82 CLIL students with 56 non-CLIL students by means of the 2000-frequency and academic bands of the receptive VLT. Participants were clustered based on both the type of instruction (CLIL/Non-CLIL) and their number of hours of exposure to make four groups: CLIL 1 ($n = 23$; 3000 hours), CLIL 2 ($n = 25$; 2400 hours), CLIL 3 ($n = 34$; 2000 hours) and non-CLIL ($n = 56$; 1200 hours). Results showed a statistically

significant difference both at the 2K band and the academic vocabulary band in favour of CLIL learners. In addition, it was found that the three CLIL subgroups did not present statistically significant differences in their receptive vocabulary, despite having different lengths of exposure. This leads the authors to suggest that it is in fact the educational context rather than the amount of input which benefits the CLIL students in terms of their vocabulary growth. Finally, Arribas (2016) compares 73 CLIL and 19 non-CLIL students in 10th grade using the 2000- and 3000-frequency bands of the receptive VLT. The number of hours of CLIL received by those in the CLIL group varied greatly, with some having received fewer than 15 hours, and others receiving between 60 to 90 hours. Results showed that although the CLIL group had a higher score in the receptive vocabulary tests, the differences were not statistically significant. The author attributes this finding to the irregular implementation of CLIL in the school at hand.

In addition to the above research in Spain, other notable studies concerning receptive vocabulary in CLIL have been carried out by Xanthou (2010) in Greece, Merikivi and Pietilä (2014) in Finland, Gierlinger and Wagner (2016) in Austria, and Sylvén and Ohlander (2014) in Sweden. Xanthou (2010) uses an experimental pretest-posttest approach to compare 60 6th grade students, divided into three groups: one experimental group of 24 students who were taught geography through L2 English for five 40-minute lessons (CLIL group), a second group of 15 students who were provided with the L1 equivalents of the target words throughout five EFL lessons (word-list group), and a final group of 21 non-CLIL students. Students were tested using an L2 to L1 translation test, whereby participants provided the L1 equivalent of 100 lexical items related to the geography unit and given in the L2, both before and after the unit. Results found that while all three groups improved from the pre-tests to the post-tests, the CLIL experimental group outperformed the other two groups in the post-test, a finding which was statistically significant. Merikivi and Pietilä (2014) compared 150 CLIL and 180 non-CLIL students from both 6th grade and 9th grade, using the 2K, 3K, 5K 10K and AWL of the VLT. For both 6th and 9th grade, results again found CLIL students to outperform their non-CLIL peers. Gierlinger and Wagner (2016) assessed students from four different Austrian lower secondary schools and compared two CLIL groups ($n = 39$) with one non-CLIL group ($n = 48$) using the X-Lex in November and May of the same academic year. Results showed that while the CLIL students did not outperform the control group in terms of overall receptive vocabulary growth, they did do so at the band of the 1000 most frequent words. Frequency analysis of teachers' input revealed that the input received by

students centred mainly around these first 1000 most frequent words, which explains this finding. Lastly, Sylvén and Ohlander (2014) report on the first data collected in a longitudinal study, which assessed students using the VLT for English (2K, 3K, 5K 10K and AWL). At the start of the project, participants included 221 students in eight groups at three different schools, with a total of 137 CLIL students and 84 non-CLIL students. CLIL students were again reported to outperform the non-CLIL students, a result which was statistically significant. The authors, however, acknowledge that the CLIL classes are voluntary, and thus will evidently be opted for by students who enjoy English, are good at it and/or feel motivated to learn other school subjects through English.

As can be seen, the majority of studies investigating receptive vocabulary in CLIL have found that CLIL students have a statistically significant higher level of receptive vocabulary than non-CLIL students. However, as has been repeatedly cautioned, this may well be due to other factors, such as higher number of hours of exposure. To investigate this issue, some researchers have compared CLIL young learners with teenage non-CLIL learners with equal hours of exposure. In these studies, the older non-CLIL students are found to outperform younger CLIL students. This, as Fernández Fontecha (2015) suggests, may well be due to the difference in learners' age or cognitive levels. This suggestion is also consistent with the findings discussed above by Iglesias Diéguez and Martínez-Adrián (2017), who found that when comparing different older and younger teenage groups with equal hours of exposure, CLIL students did indeed outperform their older non-CLIL peers. Thus, what is seen is that, when comparing 4th grade CLIL and 7th grade non-CLIL students with equal hours of exposure, the older non-CLIL do better. However, when the difference in age is reduced, and 7th grade CLIL and 8th grade non-CLIL are compared, the slightly younger CLIL students outperform the non-CLIL students.

5.3.2. Productive Vocabulary in CLIL

Unlike the receptive vocabulary research discussed above, in which the majority of studies used VLT, there has been considerably more variety in the methods used to assess productive vocabulary in a CLIL context. A first group of studies, summarised in Table 5.8, have used a range of specific vocabulary tests including picture label tasks, the Lex30, the LAT, cloze tests and the PVLTL.

Table 5.8*Research Investigating Productive Vocabulary in Spanish CLIL: Vocabulary Tests*

Authors	Age group	Tests	Main Findings
Moreno Espinosa (2009)	Young learners (6 th grade) 65 CLIL + 65 non- CLIL	Lex30	No statistically significant differences in results of the two groups.
Jiménez Catalán & Ojeda Alba (2009b)	Young learners (6 th grade) 42 CLIL + 44 non- CLIL	LAT	The non-CLIL group did significantly better on a language placement test than the CLIL group and produced a higher number of words in the LAT than the CLIL group (no statistically significant difference).
Juan-Garau et al. (2015)	Teenagers (8 th to 10 th grade) 70 CLIL + 35 non- CLIL	Cloze Test Fill-in-the- blank tense- and-aspect test	No statistically significant difference was found between the two groups at T1; CLIL students performed significantly better in the cloze test at T2, and in both the cloze test and tense-and-aspect test at T3 and T4; CLIL students improved at a significantly significant faster pace than their non-CLIL peers.
Canga Alonso & Arribas García (2015)	Teenagers (10 th grade) CLIL + non-CLIL	PVLT	Statistically significant differences in the productive vocabulary of the two groups, to the advantage of the CLIL students.

Table 5.8 (continued)

Tragant et al. (2016)	Young learners (3 rd grade) 22 CLIL/non-CLIL	Picture label task	Students were exposed to a greater number of words and more abstract and technical vocabulary in the CLIL materials, though significant progress was made in both contexts.
Jiménez Catalán & Agustín Llach (2017)	Teenagers (8 th grade) 24 CLIL + 26 non-CLIL Teenagers (10 th grade) 19 non-CLIL	LAT	The 8 th grade CLIL group outperformed both their non-CLIL peers and also the 10 th grade non-CLIL group.

Regarding young learners, Tragant, Marsol, Serrano and Llanes (2016) analyse the productive vocabulary of a small group of students ($n = 22$) at the beginning and end of a term of EFL and then at the beginning and end of term of CLIL instruction in science. Students were assessed using a specifically created picture label task which included the target vocabulary from the units which were covered in class. Textbook analysis, classroom observations, teacher interviews and questionnaires were also used. Results found that while significant progress was made in both contexts, students were exposed to a greater number of words and more abstract and technical vocabulary in the CLIL materials. Attitudinal data also suggests that learning English through science was a more challenging experience than learning in a typical EFL class. Moreno Espinosa (2009) compared 130 female CLIL and non-CLIL students (the same informants as in Jiménez Catalán and Ruiz de Zarobe, 2009, discussed above) using the Lex30 to assess productive vocabulary breadth. Results show that, regardless of the teaching approach, there were no statistically significant differences in results of the two groups. Finally, as discussed in Section 2.5.3, Jiménez Catalán and Ojeda Alba (2009b) compared CLIL ($n = 42$) and non-CLIL ($n = 44$) students in 6th grade, using a LAT. Results indicated that the non-CLIL group did significantly better on a language placement test than the CLIL group. They also produced a higher number of words in the LAT than the CLIL group; however, the difference was not statistically significant.

With regards to teenagers, Juan-Garau, Prieto-Arranz and Salazar-Noguera (2015) analysed the lexico-grammatical development of 70 CLIL and 35 non-CLIL students using a cloze test and a fill-in-the-blank tense-and-aspect test. Participants were assessed on four occasions over a three-year period, while the students were in 8th, 9th and 10th grade of compulsory secondary education. Results found that although no statistically significant difference was found between the two groups at T1, CLIL students performed significantly better in the cloze test at T2, and in both the cloze test and tense-and-aspect test at T3 and T4. In addition, while both groups significantly improved their lexical and grammatical ability over the three-year period, CLIL students were found to improve at a significantly significant faster pace than their non-CLIL peers. Jiménez Catalán and Agustín Llach (2017), in order to deal with the issues regarding CLIL and time of exposure mentioned above, compared teenage CLIL and non-CLIL learners after an equal number of hours of exposure to English by means of a LAT. Participants included 24 CLIL students and 26 non-CLIL students in 8th grade, as well 19 non-CLIL students in 10th grade with an equal number of hours of exposure as the 8th grade CLIL group. Results showed that the 8th grade CLIL group not only outperformed their non-CLIL peers, but also the 10th grade non-CLIL group. Canga Alonso and Arribas García (2015) compared 101 10th grade CLIL ($n = 73$, 1,079-1,109 hours of exposure) and non-CLIL ($n = 38$, 1,049 hours of exposure) students using the 2000-word parallel version of the PVLТ. Results found that there were statistically significant differences in the productive vocabulary of the two groups, to the advantage of the CLIL students.

In addition to these studies, which use specific vocabulary tests, a second group of studies analyse written compositions in terms of their productive vocabulary. These studies analyse written composition with varying focuses, such as TTRs, lexical reiteration/variation/transfer/inventions/complexity/profiles/phrases, themes, word frequency and L1 influence. This research is outlined in Table 5.9 and discussed in turn below.

Table 5.9*Research Investigating Productive Vocabulary in Spanish CLIL: Written Compositions*

Authors	Age group	Focus	Main Findings
Jiménez Catalán et al. (2006)	*Young learners (6 th grade) 65 CLIL + 65 non- CLIL	TTRs	CLIL students produced writing with greater lexical richness, (fewer tokens and types than the non-CLIL group but a higher TTR; they also produced a higher number of lexical verbs.
Agustín Llach & Jiménez Catalán (2007)		Lexical reiteration Lexical variation	Groups performed similarly (frequent word repetition): advantages for CLIL students with regard to lexical variation, language level, and use of antonyms and general nouns.
Agustín Llach (2009)		Language proficiency Lexical transfer	CLIL advantage, with non-CLIL learners producing significantly more lexical transfer errors than the CLIL learners.
Ojeda Alba (2009)		Themes Specific vocabulary	Both groups produced equivalent number of tokens and types in the field analysed, however, a number of diversities were found, particularly in the lexical fields of animals, food and personality traits.

Table 5.9 (continued)

Celaya & Ruiz de Zarobe (2010)	Teenagers (7 th grade) 22 CLIL + 20 non- CLIL Teenagers (10 th grade) 34 CLIL + 18 non- CLIL	Borrowings Lexical inventions	Clear effect of type of instruction on borrowings, with CLIL students producing fewer borrowings than their non-CLIL peers.
Ruiz de Zarobe & Celaya (2010)	Teenagers (9 th grade) 47 CLIL Teenagers (10 th grade) 26 CLIL + 18 non- CLIL	Linguistic competence	Positive relationship between the amount of time in CLIL and the linguistic outcomes in written production; differences are more apparent in terms of content, vocabulary and language use.
Navés (2011)	Young learners (5 th grade) 72 CLIL + 110 non- CLIL Teenagers (7 th grade) 23 CLIL + 213 non- CLIL Teenagers (9 th grade) 67 CLIL + 67 non- CLIL Teenagers (10 th -11 th grade) 38 CLIL + 55 non- CLIL	Lexical complexity	CLIL learners outperformed non-CLIL learners in terms of lexical complexity; CLIL learners' range of vocabulary was as good as students two grades above.

Table 5.9 (continued)

Agustín Llach (2014)	Young learners (4 th grade) 72 CLIL	Lexical profiles Lexical transfer	Lexical profiles were characteristic for the age group, with high frequency words being the most common; frequent instances of phonetic rendering; little L1 influence for borrowings and reflexifications.
Jiménez Catalán & Fernández Fontecha (2015)	*Young learners (6 th grade) 65 CLIL + 65 non-CLIL	Lexical phrases	Statistically significant differences in favour of the CLIL group in terms of language level, but not in terms of the production of types of lexical phrases.
Agustín Llach (2016)	Young learners (4 th to 6 th grade) 68 CLIL + 61 non-CLIL	Word frequency, L1 influence in the form of borrowings and lexical creations	CLIL learners performed slightly better than their non-CLIL peers along the three years, though there was no statistically significant difference.
Agustín Llach (2017)	Young learners (5 th grade) 72 CLIL + 68 non-CLIL	Frequency bands of words used, word origin, L1 influence in lexical production, and phonetic spelling, and learners' vocabulary sizes with the VLT	Both groups produced very similar results.

Note. *Same participants

As can be seen, Agustín Llach, in particular, has contributed a number of studies which investigate the productive vocabulary of young learners by means of written compositions. Data for these studies were collected by means of a written composition, consisting of a letter to a prospective host family in England in which students had to introduce themselves and talk about their family, school, hometown, hobbies, and any other thing the host family may be interested in. Agustín Llach (2014) presents the results of the compositions written by 72 4th grade CLIL students, analysed in terms of their lexical profiles and lexical transfer. Lexical profiles were found to be characteristic for the age group, with high frequency words being the most common. Participants also produced frequent instances of phonetic rendering, however, little L1 influence for borrowings and reflexifications. Agustín Llach (2017) presents a comparison of the compositions written by these learners in 5th grade with 68 non-CLIL 5th grade students, analysing them in terms of the frequency bands of words used, word origin, L1 influence in lexical production, and phonetic spelling, and learners' vocabulary sizes with the VLT. Despite the higher number of hours of exposure on the part of the CLIL students, results indicate that both groups produced very similar results. Agustín Llach (2016) then presents a longitudinal study which follows 68 of the CLIL and 61 of the non-CLIL students over the course of three years, from 4th to 6th grade, focusing on word frequency and L1 influence in the form of borrowings and lexical creations. Results showed that although the CLIL learners performed slightly better than their non-CLIL peers along the three years, the difference was not statistically significant. Using the same instrument, Jiménez Catalán, Ruiz de Zarobe and Cenoz Iragui (2006), Agustín Llach and Jiménez Catalán (2007), Agustín Llach (2009), Ojeda Alba (2009) and Jiménez Catalán and Fernández Fontecha (2015) analysed the compositions of 130 female CLIL and non-CLIL 6th grade students (the same informants as in Jiménez Catalán and Ruiz de Zarobe, 2009, discussed above), each with a different focus. Jiménez Catalán et al.'s study (2006), which focused on the number of tokens and types, found that CLIL students produced writing with greater lexical richness, as although they produced fewer tokens and types than the non-CLIL group, their TTR was higher. In addition, the CLIL group produced a higher number of lexical verbs, indicative of higher lexical sophistication and language level. Agustín Llach and Jiménez Catalán (2007) discuss an analysis of the writings in terms of lexical reiteration and variation. Concerning lexical reiteration, groups performed similarly, resorting often to word repetition. However, advantages were revealed in the CLIL students with regard to lexical variation, language level, and use of antonyms and general

nouns. Agustín Llach (2009), who focuses on language proficiency and lexical transfer, indicates a CLIL advantage, with non-CLIL learners producing significantly more lexical transfer errors than the CLIL learners. Ojeda Alba (2009) presents a descriptive analysis of the specific themes and vocabulary used in the compositions. Results indicate similar trends in each group, with both producing an equivalent number of tokens and types in the field analysed, however, a number of diversities are also highlighted, particularly in the lexical fields of animals, food and personality traits. Finally, Jiménez Catalán and Fernández Fontecha (2015) analysed the lexical phrases in the compositions. While results showed statistically significant differences in favour of the CLIL group in terms of language level, this was not the case for the production of types of lexical phrases.

Assessing both young learners and teenagers, Navés (2011) analysed the writing proficiency of CLIL and non-CLIL students in terms of lexical complexity, as well as accuracy, fluency and syntactic complexity. Participants included learners from 5th grade (72 CLIL and 110 non-CLIL), 7th grade (23 CLIL and 213 non-CLIL), 9th grade (67 CLIL and 67 non-CLIL) and 10th-11th grade (38 CLIL and 55 non-CLIL). Results showed not only that CLIL learners outperformed non-CLIL learners in terms of lexical complexity, but also that CLIL learners' range of vocabulary was as good as the students two grades above.

Focusing specifically on teenagers, Celaya and Ruiz de Zarobe (2010) and Ruiz de Zarobe and Celaya (2010) offer a comparison of secondary CLIL and non-CLIL students, drawing from data selected from two large studies in the Basque country and Catalonia. While students in both locations completed a written composition, there were some differences: in Catalonia students were given 15 minutes to write about the topic "My life: past, present and future expectation" whereas in the Basque Country students were given 20 minutes to complete the task "Write a letter to a host family". The participants in Celaya and Ruiz de Zarobe (2010) were 7th grade CLIL ($n = 22$) and non-CLIL ($n = 20$), and 10th grade CLIL ($n = 34$) and non-CLIL ($n = 18$) students. Compositions were analysed in terms of borrowings and lexical inventions. In particular, findings show a clear effect of type of instruction on borrowings, with CLIL students producing fewer borrowings than their non-CLIL peers. The participants in Ruiz de Zarobe and Celaya (2010) were 9th grade CLIL ($n = 47$, 23 from Catalonia and 24 from the Basque Country) and 10th grade CLIL ($n = 26$, from Catalonia) and non-CLIL ($n = 18$, from the Basque Country) students. The study offers a comparison of the groups' linguistic competence, assessing the compositions in terms of content, organisation,

vocabulary, language use and mechanics. Results show a positive relationship between the amount of time in CLIL and the linguistic outcomes in written production, and in particular reveal that the differences are more apparent in terms of content, vocabulary and language use.

In addition to the research above carried out in a Spanish context, notable research on productive vocabulary in CLIL includes that by Sylvén (2004) and Olsson (2015) in Sweden, Merikivi and Pietilä (2014) in Finland, and Baten et al. (2020) in the Netherlands. Firstly, Sylvén (2004) compared both the receptive and productive vocabulary of CLIL ($n = 99$) and non-CLIL ($n = 264$) students at the beginning and end of their first and second year of upper secondary education. Although CLIL students were found to outperform their non-CLIL peers, this was found to be so even from the first data collection. However, while both groups improved significantly over time, the CLIL group did so to a greater extent. The author highlights that the superior lexical development of the CLIL group is clear, but that it may not be attributable to the CLIL approach alone. Olsson (2015) compared academic vocabulary use among CLIL ($n = 146$) and non-CLIL ($n = 84$) students, again in Swedish upper secondary school. As in Sylvén (2004), results showed an advantage on the part of CLIL students even at the beginning of CLIL education, with CLIL students using academic vocabulary to a greater extent than their non-CLIL peers. However, their use of academic vocabulary was not found to progress more than the non-CLIL students. In addition to the receptive tests outlined in Section 5.3.1, Merikivi and Pietilä (2014) also analysed the productive vocabulary of their 6th grade and 9th grade CLIL and non-CLIL participants using the PVLТ. As in the receptive tests, and contrary to the results found in the aforementioned studies investigating the productive vocabulary of young learners in CLIL, CLIL pupils were found to produce a statistically significant higher number of words than their non-CLIL peers. In addition, in line with previous research, students' receptive vocabulary was found to be a great deal larger than their productive vocabulary. Finally, of particular relevance to the project at hand is the study by Baten et al. (2020), given that it assesses the vocabulary of students enrolled in trilingual CLIL who study both English and French as an L2. The participants include 75 8th grade L1 speakers of Dutch, who receive both content and language instruction in the two TLs: two hours of history in French and one hour of music in English per week; and four or five hours of French language and two hours of English language per week. Students were assessed on two occasions on both their receptive and productive vocabulary, using the VLT and PVLТ at the 2000-, 3000-, 5000- and 10000-

word levels and the academic word level. Results indicated that while the participants had larger vocabularies in English, the level of improvement between the pre- and post-tests was not different across the two languages. These findings are particularly important in trilingual CLIL research as they provide evidence that CLIL vocabulary knowledge develops in LOTEs.

In summary, contrary to the supposition that productive vocabulary is unaffected by a CLIL approach, the above studies indicate that the benefits of CLIL on productive vocabulary may depend to some degree on age. In general, no statistically significant differences are found between young CLIL and non-CLIL learners (with the exception of Merikivi and Pietilä, 2014). A number of studies do, however, suggest that on closer inspection, there may be some qualitative advantages, with CLIL students demonstrating higher lexical richness, sophistication and variation and fewer lexical transfer errors. Among adolescents, on the other hand, CLIL students have by and large been found to outstrip their non-CLIL peers. In addition, studies which have compared CLIL and non-CLIL adolescents in different grades but with equal hours of exposure have found the younger CLIL students to outperform older non-CLIL students in terms of productive vocabulary. This finding is consistent with results found in studies investigating receptive vocabulary.

5.4. Research on Motivation and CLIL

As has been discussed in detail in Chapter 3 (see [Section 3.5.3](#) for an in dept overview), research into motivation in a CLIL context has only begun to emerge in earnest in the past five years. This research has either aimed at measuring the learners' L2 motivation quantitatively or qualitatively in a monolingual or multilingual context or investigating the beliefs of different stakeholders. Firstly, regarding the former, a great deal of variation is evident in the research to date, likely due to the diverse methodological approaches used, the different learning environments in which the studies are carried out, and the varying profiles of the participants. In general, research has shown that CLIL students are more motivated than their non-CLIL peers (e.g., Lasagabaster & López Beloqui, 2015; Verspoor et al., 2015; Pfenninger, 2016; De Smet et al., 2018, 2019), however, some research has found this not to be statistically significant (e.g., Heras & Lasagabaster, 2015; Rumlich, 2016, 2017; Lasagabaster, 2017) or even find that non-CLIL students are more motivated than CLIL students (e.g., Fernández Fontecha & Canga Alonso, 2014). Secondly, regarding the latter, stakeholders' views of CLIL have

largely been positive in research to date, viewing CLIL as a motivational approach, especially when students take part in more intensive tracks (Somer & Llinares, 2018). However, concerns have been raised by parents and teachers with regard to language level and acquisition of content (Pladevall Ballester, 2015), the extent to which parents can offer extramural support in a language they are unfamiliar with (Ráez Padilla, 2018), and inadequate teacher training (Barrios & Milla Lara, 2020).

5.5. Research on Gender and CLIL

As has been discussed in detail in Chapter 4 (see [Section 4.3.3](#) for a detailed overview), although a female advantage is often found in language learning both in terms of language proficiency and language motivation, suggestions have been made that this may be dependent on the teaching context. For example, authors such as Merisuo-Strom (2007) and Gallardo-del-Puerto and Blanco-Suárez (2021) found statistically significant differences in non-CLIL students but no such difference in CLIL students in terms of attitudes and motivation, respectively. Fernández Fontecha (2014a) similarly found no statistically significant difference between male and female CLIL students in terms of motivation, whereas Fernández Fontecha (2014b) revealed no difference in terms of extrinsic motivation but statistically significant differences in terms of intrinsic motivation. Comparing CLIL and non-CLIL learners, Fernández Fontecha and Canga Alonso (2014) again found no statistically significant difference between male and female learners' motivation in CLIL; however, the same results were found in non-CLIL. Other research has shown that female students outperform male students, regardless of teaching context, both in terms of language skills (e.g., Lasagabaster, 2008) and language attitudes (e.g., Lasagabaster & Sierra, 2009). Despite this female advantage, some research has also indicated that female students may also score higher in negative motivation categories such as anxiety and self-confidence (e.g., Sylvén & Thompson, 2015). This research has found both gender and context to play a key role, given that non-CLIL females reported the highest English anxiety and lowest L2 self-confidence, followed by CLIL females, non-CLIL males, and finally CLIL males, who had the lowest anxiety and highest L2 self-confidence. Other research has indicated that potential differences may arise in different aspects of motivation, such as that by Heras and Lasagabaster (2015), which revealed that while males had higher Ought-to Self than females in CLIL, females had higher Ideal L2 Self than males in non-CLIL. In sum, although some studies have indeed found a potential levelling effect of CLIL on gender, others have found females to

outperform males in terms of language skills, attitudes and motivation, regardless of the teaching context. This may potentially be attributed to the more limited form of CLIL at hand, as opposed to immersion which has been found to have a greater effect (Lasagabaster & Sierra, 2009). As is suggested by Heras and Lasagabaster (2015), this blurring of gender differences may come about as a result of male students' interest in the content subject. If this is the case, it is essential for future research to analyse students' specific motivation and interest towards the content studied in CLIL, so as to determine whether, when studying a subject which they particularly like through the TL, there is a consequential increase in the students' language proficiency and language learning motivation.

Chapter 6: Research Questions and Hypotheses

This chapter will first outline the main rationale for carrying out this thesis, drawing from the literature provided in Chapters 2 to 5, and the resulting objectives it aims to meet. It will then indicate the thesis' research questions and hypotheses.

6.1. Research Rationale and Objectives

As has been outlined throughout the previous four chapters, the central focus of this thesis is on vocabulary, and in particular LA, L2 motivation and gender, all within a secondary CLIL learning context and with regards to not only English but also other FLs such as French. This section will provide a summary of the key research gaps in these areas that have been outlined in previous chapters.

As has been discussed in Chapter 5, CLIL teaching has progressed by leaps and bounds since the turn of the century and, following from the initial excitement of the approach, research has frantically been trying to play catch up (Pérez-Cañado, 2012). While the approach becomes more widespread each year, areas which are lacking and aspects which need to be improved continue to be acknowledged.

Firstly, regarding the language of instruction, despite the fact that the original objective behind adopting a CLIL approach was to promote multilingualism, there has been a remarkable and undeniable predominance of the use of English throughout CLIL programmes and inevitably in CLIL research (San Isidro, 2018). Dalton-Puffer et al. (2010) have consequently called for comparative research across additional LOTEs in order to provide insight into the strengths and weaknesses of CLIL language-independently. Similarly, Cenoz et al. (2014) have called for a more critical, empirical examination of CLIL in diverse contexts, focusing not just on ESL/EFL but also on other L2s which have been largely neglected. This need has yet to be properly addressed, with Merino and Lasagabaster (2018a) again highlighting the clear lack of studies dealing with the effects of three languages in CLIL.

Secondly, regarding vocabulary, research has shown that CLIL modules produce positive effects on students' learning of content-related vocabulary (Heras & Lasagabaster, 2015). Baten et al. (2020) have recently provided valuable insight into the vocabulary acquisition of Dutch students simultaneously studying English and French as L2s in a trilingual CLIL context, taking the first steps to address Merino and Lasagabaster's (2018a) call for such research above. This research indicated that participants had larger

vocabularies in English, though the level of improvement between the pre- and post-tests was not different across the two languages, and highlighted the need for further investigation of the effect of the subject matter on CLIL learners' L2 development. Likewise, and specifically with regards to LA, Canga Alonso (2017) has highlighted the need to focus on LA prompts which may be relevant to the CLIL subject, for example, investigating the prompt *Parts of the Body* with students who have learnt Natural Sciences and Biology through English. He also pointed out the necessity of including some measure of proficiency alongside the LAT, so as to determine the effect of language level on the number of words retrieved. LA research has also primarily focused on L1s or L2s, with a scarcity of studies examining LA in two L2s. One exception to this has been research by Santos Díaz (2017c), who analysed the LA of Spanish university students in Spanish (L1) and either English (L2) or French (L2). This research provided data on the number of words produced in Spanish (397.67), English (261.67) and French (221.33), as well as the percentage of non-shared words in the three languages: 42.44% in Spanish, 44.76% in English and 50.09% in French. While this research offers a glimpse of the differences between the students' L1 and one L2, the data in English and French is evidently not directly comparable, given that it comes from different groups of students. There is thus clearly a need for comparative research into students' LA in their various L2s to determine any similarities or differences in each one, as well as LA research in CLIL contexts which specifically targets vocabulary acquired in content subjects.

Thirdly, regarding motivation, Navarro and García Jiménez (2018, p. 87) have pointed out that the role of motivation is now understood to be "an unequivocally important factor for the learning of an L2, and it plays a more important role in CLIL than in non-CLIL settings". Furthermore, regarding motivation in different L2s, Dörnyei and Ushioda (2013) highlight how the spread of English as a global language and international lingua franca can lead to a qualitative difference between the motivation for learning English compared with other languages, given that English is increasingly being seen as a basic educational skill, imperative to professional advancement. In a Belgian context, De Smet et al. (2018, 2019) have carried out research with French-native Belgian students studying either English or Dutch in primary and secondary CLIL contexts and showed that students in English CLIL reported significantly less anxiety, more positive attitudes and higher motivation than those in Dutch CLIL, indicating a clear difference in CLIL motivation depending on the language of study. In Spain, while Geoghegan (2018) found statistically significant differences between SA university students focusing on studying

English as compared with French or German, there is lack of research on this issue in a secondary CLIL context. Furthermore, these studies compare students' motivation towards one language, rather than towards both TLs simultaneously. It is thus suggested that research must investigate the differences between CEIL as opposed to CLIL among the same cohort in secondary education, particularly with regards to language learning motivation and its effect on vocabulary acquisition. This would also meet Oakes and Howard's (2019) call for not only motivation research on LOTEs, but for better integration between research on LOTEs and EFL.

Finally, regarding gender, research investigating LA in a Spanish context has largely shown a female advantage, whereby female students in 6th to 9th grade have produced a statistically significant higher number of words in the LAT as compared with their male peers. However, more recent research by Jiménez Catalán and Canga Alonso (2019) with 12th grade students, found no such difference between male and female learners, indicating a potential age-related effect. Given the lack of studies with Spanish secondary students in 10th and 11th grade, there is a clear need to address the LA of this age group, so as to determine whether there is a difference between gender in late primary and early secondary, which progressively diminishes towards the end of secondary education. Research into gender and motivation has also largely found that female students generally report higher motivation or positive attitudes (e.g., Fernández Fontecha, 2010; Fernández Fontecha, 2014c for intrinsic motivation; Lasagabaster & Sierra, 2009; Lasagabaster, 2016; Merisuo-Storm, 2006; Sylvén & Thompson, 2015; Calafato & Tang, 2019). However, when investigating gender and motivation in CLIL, findings have also indicated that students show clear differences depending on the teaching context (e.g., Fernández Fontecha & Canga Alonso, 2014; Lasagabaster and Sierra, 2009; Heras & Lasagabaster, 2015; Gallardo-del-Puerto & Blanco Suárez, 2021). Taking into account suggestions regarding the blurring of gender-based differences in CLIL due to male students' interest in the subject context (Heras & Lasagabaster, 2015), it is of the utmost importance that research on gender, motivation and CLIL consider students' specific motivation and interest towards the content studied in CLIL, so as to confirm whether there is a consequential increase in language proficiency and language learning motivation.

In addition to the above variables, it is also important to highlight the need to investigate these issues using a longitudinal approach. As is noted by Ortega and Ibarra-Shea (2005), there is a clear lack of longitudinal research in applied linguistics, despite

the benefits that this research would bring to the field of SLA in providing a meaningful interpretation of language learning, development, progress, change or gains. Specifically regarding CLIL, this scarcity has repeatedly been flagged, given that it prevents conclusive evidence for the added value of CLIL in L2 learning (Pérez Cañado, 2018a; Goris et al., 2019). It is thus clear that, whenever possible, SLA research needs to take a longitudinal approach in order to appropriately analyse language learning in this context.

In order to address these various issues, the current study focuses on male and female secondary school students who simultaneously take both CEIL (content through English) and CLIL (content through another language: French) and investigates the LA and the language learning motivation of male and female students by means of pre- and post-tests. The main objectives of the project are as follows:

1. to determine whether there are quantitative differences between the learners' lexical availability in English and French.
2. to ascertain whether there are qualitative differences between the learners' lexical availability in English and French with regards to a) the frequency of first word responses for each prompt, b) the most and least productive prompts, and c) the lexical sophistication of each language based on the non-shared words of participants as well as d) the number of infrequent words in the production of each prompt.
3. to determine whether there is a difference between the participants' English language learning motivation as compared with their French language learning motivation.
4. to determine whether there is a relationship between the lexical availability, language learning motivation and language level of the participants in each language.
5. to ascertain whether there are quantitative differences between male and female learners' lexical availability in English and French.
6. to ascertain whether there are qualitative differences between male and female learners' lexical availability in English and French with regards to a) the frequency of first word responses for each prompt, (b) the most and least productive prompts, and c) the lexical sophistication of each language based on the non-shared words of participants as well d) as the number of infrequent words in the production of each prompt.

7. to determine whether there is a difference between the male and female participants' language learning motivation in English and French.
8. to ascertain whether there is an effect of CLIL instruction on the students' lexical availability and language learning motivation in each language.
9. to determine whether there are differences in the lexical availability, language learning motivation and language level of the participants in each language at each testing period.

6.2. Research Questions and Hypotheses

This section outlines the research questions and hypotheses for this thesis, which are organised into four key areas: Lexical Availability, Language Learning Motivation, Gender and CLIL.

6.2.1. Target Language Influence on Lexical Availability

With regard to LA in English and French, the following research question is posed:

RQ1: What are the differences in the lexical availability of secondary school CLIL students in English as compared to French?

RQ1.1: Are there quantitative differences in the words retrieved by participants in LATs in English and French at each testing period?

RQ1.2: Are there qualitative differences in the words retrieved by participants in LATs in English as compared to French at each testing period with regards to a) the frequency of first word responses for each prompt, (b) the most and least productive prompts, and c) the lexical sophistication of each language based on the non-shared words of participants as well d) as the number of infrequent words in the production of each prompt?

While there is limited research comparing LA in different FLs, the few studies that have been carried out have indicated that participants produce a higher number of words in English as compared with other FLs (Santos Díaz, 2017c; Šifrar Kalan, 2014). The role of language level has also been highlighted, with advanced learners and those with greater exposure to the language producing the highest number of words (van Ginkel & van der Linden, 1996, as cited in Schmitt, 2000). It is thus hypothesised firstly that, given the predominant role of EFL in Spain and the higher number of hours of exposure that students generally receive, participants will retrieve a higher number of words in English

as compared with French. Regarding quantitative differences at each testing point, it is expected that older, more advanced students will produce a higher number of words than younger students in each language, in keeping with findings from previous research (Jiménez Catalán & Fitzpatrick, 2014; Agustín Llach & Fernández Fontecha, 2014). Secondly, regarding the qualitative aspects of LA, it is hypothesised that in the case of the frequency of first word responses for each prompt and the most and least productive prompts, similarities will be observed in both languages in general prompts, in keeping with Santos Díaz's (2017c) findings for the latter. However, given the research which has shown CLIL modules to have positive effects on students' learning of content-related vocabulary (Dalton-Puffer, 2008; Heras & Lasagabaster, 2015), it is also hypothesised that when a prompt is related to a subject which is studied through the medium of one of the TLs, this prompt will be more productive in that language. Regarding lexical sophistication, in keeping with findings by Šifrar Kalan (2014), it is expected that more non-shared words will be found in the least productive prompts. In addition, based on the findings from Santos Díaz (2017c), it is expected that participants will produce fewer non-shared words in English than in French. On the basis of Jiménez Catalán and Fernández Fontecha's (2019) findings comparing the LA of L2 and L3 students, it is also expected that participants will produce a higher percentage of infrequent words in the most productive prompt in their L2 than in their L3. Finally, regarding qualitative differences at each testing period, it is hypothesised that while the frequency of first word responses for each prompt and the most and least productive prompts will be similar at different grades for some prompts, other prompts may reveal differences, as in Jiménez Catalán and Fitzpatrick (2014). In particular, the most and least productive prompts are expected to vary in the case of content-relevant prompts, when subjects studied differ in each grade. In keeping with findings on lexical sophistication from Jiménez Catalán and Fitzpatrick (2014), it is also expected that older students will have incorporated more words at the 1K and 2K levels, as opposed to off-list words, in each language.

6.2.2. Target Language Influence on Language Learning Motivation

With regard to language learning motivation in English and French, the following research question is posed:

RQ2: What is the role of motivation in the language learning of secondary school CLIL students?

RQ2.1: Is there a quantitative difference between the participants' language learning motivation towards English as compared to French at each testing period?

RQ2.2: Is there a relationship between the participants' lexical availability, language level and their language learning motivation in each language at each testing period?

Previous research investigating motivation towards L2 English as opposed to other TLs has generally indicated that students report more positive attitudes and higher motivation towards English than to the other language (De Smet et al., 2018, 2019; Geoghegan, 2018). It is thus expected that students at all testing periods will report higher language learning motivation towards English than to French. Regarding the relationship between the participants' LA and language learning motivation, previous research has highlighted significant correlations between productive vocabulary and motivation (de la Maya Retamar, 2016) and specifically between LA and motivation (Fernández Fontecha, 2010). Consequently, it is hypothesized that participants who report higher language learning motivation will also produce a higher number of tokens on the LAT. In addition, given previous findings on the relationship between language proficiency and language learning motivation in a secondary CLIL context (Navarro Pablo, 2018), it is in turn expected that higher performing students will also be those who report higher motivation. This is expected to be particularly important in the case of French, given suggestions by Dörnyei and Al-Hoorie (2017) that LOTE learners generally obtain a higher language level for highly specific and personalised reasons.

6.2.3. Gender

With regard to gender, the following research question is posed:

RQ3: What are the gender differences of secondary school CLIL students of English and French with regards to lexical availability and motivation?

RQ3.1: Are there quantitative differences in the words retrieved by male and female students in LATs in English and French at each testing period?

RQ3.1.1. Are there quantitative differences between the words retrieved by male and female participants across languages, in English as compared to French, at each testing period?

RQ3.1.2. Are there quantitative differences between the words retrieved by male and female participants within languages, in English and in French, at each testing period?

RQ3.2: Are there qualitative differences in the words retrieved by male and female students in a LAT in English as compared to French at each testing period with regards to a) the frequency of first word responses for each prompt, b) the most and least productive prompts, and c) the lexical sophistication of each language based on the non-shared words of participants as well as d) the number of infrequent words in the production of each prompt?

RQ3.3: Are there quantitative differences between male and female students with regards to language learning motivation in English and French at each testing period?

RQ3.3.1. Are there quantitative differences between male and female participants with regards to language learning motivation across languages, in English as compared to French, at each testing period?

RQ3.3.2. Are there quantitative differences between male and female participants with regards to language learning motivation within languages, in English and in French, at each testing period?

It is firstly expected that both male and female students will produce a higher number of words in English as compared to French, as in RQ1. Secondly, based on the findings of previous research investigating LA and gender, it is predicted that female students will produce a higher number of words than male students (Jiménez Catalán & Ojeda Alba, 2009a; Fernández Fontecha, 2010; Agustín Llach & Fernández Fontecha, 2014). However, given findings by Jiménez Catalán and Canga Alonso (2019) which found no gender difference in 12th grade students, it is suggested that gender difference may become less pronounced as students get older. Regarding qualitative differences, similarities are expected across genders in the most and least productive prompts, as in Agustín Llach and Fernández Fontecha (2014). However, in cases where male and female students study different content subjects, it is hypothesised that the number of words produced in content-relevant prompts will vary across genders, with those receiving more prompt-related vocabulary exposure producing a higher number of words in that category.

Finally, regarding language learning motivation, it is first and foremost expected that both male and female students will report higher motivation towards English as compared to French, as in RQ1, and secondly that female students will report higher motivation than male students towards each language, in keeping with previous findings (Fernández Fontecha, 2010; Fernández Fontecha, 2014c; Lasagabaster & Sierra, 2009; Lasagabaster, 2016; Merisuo-Storm, 2006; Sylvén & Thompson, 2015; Calafato & Tang, 2019). Nonetheless, given findings indicating a blurring effect of CLIL on motivation (Gallardo-del-Puerto & Blanco Suárez, 2021) and the supposition that male students may be more motivated in a CLIL context due to greater interest in the subject (Heras & Lasagabaster, 2015), it is hypothesised that male students who express greater interest in the content classes will consequently have greater language learning motivation, perhaps to the extent that they match their female peers.

6.2.4. Content and Language Integrated Learning

With regard to CLIL, the following research question is posed:

RQ4: What effect does CLIL instruction have on the language learning of secondary school students of English and French?

RQ4.1: Is there a quantitative difference in the language level, lexical availability and motivation of students taking different CLIL classes at each testing period?

RQ4.2: Is there a qualitative difference in the words produced in English by students taking different CLIL classes at each testing period with regards to a) the frequency of first word responses for each content-related prompt, b) the most and least productive content-related prompts, and c) the lexical sophistication of each language based on the non-shared words of participants as well as d) the number of infrequent words in the production of each content-related prompt?

RQ4.3: Is there a quantitative difference in the number of words produced by students in English and French when the prompt is related to a CLIL class taken in either English or French at each testing period?

RQ4.4: Is there a qualitative difference in the words produced by students in English and French when the prompt is related to a CLIL class taken in either English or French at each testing period with regards to a) the frequency of first word responses for each content-related prompt, (b) the most and least productive

content-related prompts, and c) the lexical sophistication of each language based on the non-shared words of participants as well as d) the number of infrequent words in the production of each content-related prompt?

As mentioned above, it is expected that participants with a higher language level will generally produce a higher number of words of the LAT. However, as has been pointed out by Canga Alonso (2017), there is a need to include LA prompts which may be relevant to the CLIL subject alongside some measure of language proficiency, in order to determine whether there is a relationship between language level and LA in content-specific prompts. A CLIL approach has been shown to have a positive effect on students' content-related vocabulary (Dalton-Puffer, 2008; Heras & Lasagabaster, 2015). It is thus expected that when taking the content-related prompts into consideration, differences will arise in LA, both quantitatively and qualitatively, depending on the exposure students have received to this vocabulary in their CLIL classes. In terms of motivation, it is expected that students who express greater motivation towards a given CLIL subject will also produce a higher number of words in the related prompt, in keeping with research which has found a relationship between LA and motivation (Fernández Fontecha, 2010). Regarding English and French, it is expected that prompts related to CLIL classes in English will see students produce a greater number of words in the English LAT, while prompts related to French will see students produce a greater number of words in the French LAT. This is based on research on secondary Spanish CLIL which has found CLIL groups to outperform non-CLIL groups (Jiménez Catalán & Agustín Llach, 2017). In the case of general prompts, it is expected that there will be no difference between the students enrolled in different CLIL classes. Given the greater exposure to content-related vocabulary, qualitative differences are also expected to be found in the case of content-relevant prompts, with prompts related to English CLIL classes resulting in a higher lexical level in English and prompts related to French CLIL classes resulting in a higher lexical level in French. Greater lexical sophistication is also expected by students in prompts which are related to their CLIL classes in each language.

The following section will now explain in detail how we intend to answer these research questions.

Chapter 7: Methodology

In this section, the methodology of the study will be described, outlining its design, the participants, the instruments, the data collection process, and finally the data analysis procedure.

7.1. Research Approach and Design

This section outlines the research approach and design of the study, which took place at two points with a period of a year in between. The first data collection took place in late January / early February of the first year of the project while the first groups of students were in 10th grade. While the second data collection was originally intended to take place after five months, that is, at the end of the term while the students were still in 10th grade, the ongoing pandemic made this impossible and so the schools were contacted to rearrange the second data collection. An unavoidable consequence of this change was that the subjects which these students studied in 10th grade were no longer the same, given differences in CLIL instruction across different grades. Of particular relevance to this project was the fact that the subjects which the students studied from 7th to 10th grade in English (e.g., Physical Education, Economics, Physics and Chemistry or Biology) and in French (e.g., History and Geography) were generally studied through Spanish in 11th grade. The only exception to this was the male group's economics students, who continued to study this subject through English in 11th grade. Given that the original participants were no longer enrolled in CLIL classes, the decision was made to introduce a cross-sectional approach into the study, comparing two further groups from 9th grade with the first two groups, as outlined in Table 7.1.

Table 7.1

Longitudinal and Cross-Sectional Approach of the Study

	Longitudinal	
	Year 1	Year 2
9 th grade (Groups 1 + 2)	10 th grade → 11 th grade (Groups 3 + 4)	
Cross-sectional		
	Year 2	

In this way, performance on the tests could be evaluated cross-sectionally, comparing students in 9th grade (Group 1 and Group 2) with students with an extra year's CLIL

experience in 10th grade (Group 3 and Group 4), and longitudinally, comparing the data collected from Groups 3 and 4 at the first and second data collections, that is, when they were in 10th and 11th grade. As a result of this addition, a total of twelve data collection sessions were carried out, as outlined below.

Table 7.2*Design of the Study*

	Year 1 January (Group 3) February (Group 4)		Year 2 January (all groups)	
	English	French	English	French
			9 th grade	
Group 1			1. C-test 2. LAT 3. MFQ	4. C-test 5. LAT 6. MFQ
Group 2			1. C-test 2. LAT 3. MFQ	4. C-test 5. LAT 6. MFQ
	10 th grade		11 th grade	
Group 3	1. C-test 2. LAT 3. MFQ	4. C-test 5. LAT 6. MFQ	1. C-test 2. LAT 1. MFQ	3. C-test 4. LAT 5. MFQ
Group 4	1. C-test 2. LAT 3. MFQ	4. C-test 5. LAT 6. MFQ	2. C-test 3. LAT 4. MFQ	5. C-test 6. LAT 7. MFQ

As shown in Table 7.2, at each data collection session each group was given three tests dealing with English (50 minutes) and three tests dealing with French (50 minutes), for a total of six tests. The tests included a C-test, a lexical availability task (LAT) and a questionnaire (MFQ) in each language. In the case of Group 1 and Group 2, these tests were administered once, in Year 2. In the case of Group 3 and Group 4, these tests were administered both in Year 1 and again in Year 2 after a period of 12 months. Following research from Kremmel and Schmitt (2018), who advise that the same version of tests should be used twice provided there is a minimum of one month between testing, Group 3 and Group 4 took the same tests at both data collections.

7.2. Participants

The participants in this study were Spanish native speakers ($N = 91$) in 9th, 10th and 11th grade in two semi-private sister schools, one for boys and one for girls, which place a focus on plurilingual education, conducting classes in Spanish, English, and French. As part of their plurilingualism project, the schools offer this linguistic immersion from early childhood education and throughout primary education and obligatory secondary education, that is, from age 3 until approximately age 16. The languages are used as a vehicular language, with each language taking up a third of the students' school day. As part of the curriculum, students are expected to reach a B2 or C1 level in their TLs by the time they finish school. Students may also opt to take on an additional fourth language. Participants were divided into groups according to grade and gender: Group 1 (9th grade male), Group 2 (9th grade female), Group 3 (10th-11th grade male) and Group 4 (10th-11th grade female) (Table 7.3).

Table 7.3

Participants in the Study

	9 th grade	10 th grade	11 th grade	Total
Male	16 (En) / 19 (Fr)	14 (En) / 12 (Fr)		39
		1	5 (En) / 3 (Fr)	
Female	23	18		52
		8	3	
Total	42	32		91
		9	8	

As shown above, while participants in 10th and 11th grade were largely the same, there were a total of 17 students (6 male and 11 female) who took part only in one data collection. This was due largely to the fact that the transition from 10th to 11th grade also entails a changeover in school stage (from obligatory secondary education to high school), and so a number of students no longer attended the school in 11th grade while other students enrolled in 11th grade. In other cases, this was due to absences on the day of one data collection. In addition, in the male groups, some students took part in the tests for only one language. In the 9th grade group, all male students took part in the French tests ($n = 19$), while three students later decided not to participate in the English tests. In the

11th grade group, although French had been an obligatory subject the previous year, in 11th grade it became optional, and so there were four students who no longer studied French and consequently did not participate in the French tests.

In addition to EFL and FFL, CLIL subjects offered in the schools included a science subject (Biology and/or Physics and Chemistry), Economics, Physical Education and Technology through English; and Geography and History through French (Table 7.4).

Table 7.4

Participants' CLIL Subjects

		9 th grade	10 th grade	11 th grade
Male	En	1. Physics & Chemistry 2. Biology 3. Physical Education 4. Technology	1. Physics & Chemistry ($n = 10$) or Economics ($n = 5$) 2. Physical Education	1. Economics ($n = 10$)
	Fr	1. Geography & History	1. Geography & History	N/A
Female	En	1. Physics & Chemistry 2. Biology 3. Physical Education	1. Biology ($n = 16$) or Economics ($n = 10$) 2. Physical Education	N/A
	Fr	1. Geography & History	1. Geography & History	N/A

Note. Numbers in parentheses indicate the number of students in 10th and 11th grade who have chosen to study either a science subject or economics. In all other cases, all students from each group study in the indicated subject.

It should be noted, although the schools' plurilingualism project dictates that one third of the school day be dedicated to each language, there appears to be a focus on English over French, with students taking English language classes and up to four CLIL classes in English, compared to taking French language classes and just one CLIL class in French. As shown above, these subjects varied depending on grade and school. While in 9th grade, students took both physics and chemistry and biology in English, in 10th grade male students had the option to study physics and chemistry while female students had the

option to study biology. In addition, in 10th grade students chose either to study the science subject or to take up economics, which had not been studied previously. Thus, in 10th grade, students formed two groups: those who studied a science subject and those who studied economics. In both 9th and 10th grade, all students studied physical education through English and geography and history through French. Finally, in 11th grade, the majority of subjects ceased to be taught through an FL. The one exception to this was economics, which continued to be studied through English by the ten male students who had chosen to study it. In order to address the effect of the differences in the CLIL subjects, further subgroups in 10th and 11th grade were included, outlined in Table 7.5.

Table 7.5

Subgroups of Students Taking Different CLIL subjects in 10th and 11th Grade

	10 th Grade	11 th Grade	10 th + 11 th Grade
Science	26 (English)	21 (Spanish)	18
Economics	15 (English)	19 (9 Spanish and 10 English)	10

These groups were made up of two groups in 10th grade: those taking a science subject in English (Physics and Chemistry, or Biology) ($n = 26$) and those taking economics in English ($n = 15$); and two groups in 11th grade: those taking a science subject in Spanish (Physics and Chemistry, or Biology) ($n = 21$) and those taking economics in English or Spanish ($n = 19$; 10 through English and 9 through Spanish). While in the majority of cases, students who had studied a science subject or economics in 10th grade continued to study this subject in 11th grade, in some cases, students opted to change subjects in 11th grade, following the transition into high school. This entailed that there were two groups from 10th to 11th grade who studied the same CLIL subject in both grades: science ($n = 18$) and economics ($n = 10$). Given that students in 9th grade all studied the same CLIL subjects, no subgroups were made in this grade.

The participants were selected based on homogeneous sampling (Dörnyei, 2007) given that they attended schools where both English and French were used as vehicular languages to teach other content subjects (Physics and Chemistry, Biology, Economics, Physical Education, Technology, and Geography and History). While many other schools in the region have incorporated CLIL in English into the curriculum, the two

abovementioned schools were the only ones in the region which also conduct CLIL in French.

7.3. Data collection

This section outlines the steps taken in the data collection, first explaining the instruments that were used and then detailing the procedure that was taken.

7.3.1. Instruments

The instruments for this study consisted of six tests: a C-test, LAT and a questionnaire, each in English and in French.

7.3.1.1. C-tests. Given its attested reliability, as outlined in the literature review, the C-test was used in this study as an independent proficiency measure. The C-tests for each language were selected from those which had been tested and used in previous studies to further ensure their validity.

The English C-test (Appendix A) was adopted from Daller and Phelan (2006). Four texts were selected, with a total of 80 gaps (20 gaps per text). Each text had been taken from an online news website. The French C-test (Appendix B) was adopted from McManus (2011), based on a version used by the Learner Language Project research team. Four texts were selected with a total of 112 gaps (between 19 and 35 gaps per text), all taken from a published newspaper article. So as to have a degree of consistency between the two C-tests with regards to length and time, gaps in the final sentences of two texts were removed so as to have a final version with a total of 80 gaps (between 19 and 21 gaps per text). In both C-tests, from the second sentence onwards, the second half of each second word was deleted and replaced with a blank space. The exercise was preceded by a sample text, which the researcher used to explain the task.

7.3.1.2. Lexical Availability Tasks. The second test given to participants was the LAT, again in both English and French (Appendix C and D). It was a paper-and-pencil questionnaire, whereby participants were presented with prompts and told to write down any words that came to mind. The prompts were presented on a different page with numbered lines, and participants were given two minutes for each prompt. For the current study, five semantic domains, which were the same for both English and French, were chosen (Table 7.6).

Table 7.6*Lexical Availability Task Prompts*

Lexical Availability Task English	Lexical Availability Task French
a. Animals	a. Les animaux
b. Food and drink	b. La nourriture et les boissons
c. Sport and physical activities	c. L'environnement et le climat
d. Environment and climate	d. Le sport et les activités physiques
e. Economy and money	e. L'économie et l'argent

The first two prompts, *Animals* and *Food and Drink* were taken from previous LATs carried out with a similar sample, namely Spanish L1 teenagers (Canga Alonso, 2017; Fernández Orío & Jiménez Catalán, 2015). These two prompts have notably been highlighted as being extremely productive, likely due to the high level of exposure to these semantic fields (Fernández Orío & Jiménez Catalán, 2015). The other three categories were related to the students' content classes: *Environment and Climate* (Geography and History, studied through French), and *Sport and Physical activities* and *Economy and Money* (Physical Education and Economics, studied through English). While these categories were catered specifically towards the students' content subjects, similar categories have also been investigated in previous studies, such as *Sports and Hobbies* (Agustín Llach & Fontecha, 2014) and *The Environment* and *The Economy* (Neilson Para, 2016).

7.3.1.3. Questionnaires. Two questionnaires were created for this study. While the two questionnaires, both written in Spanish, were very similar, one was designed to investigate the participants' motivation towards learning English (Appendix E) while the other was designed to investigate their motivation towards learning French (Appendix F). Each questionnaire consisted of two parts. The first part had eight questions and dealt with personal information such as the participant's age, gender, nationality, and language learning background. The second part consisted of the MFQ which consisted of around 55 questions for the English MFQ and 51 questions for the French MFQ. The questions followed a five-level Likert scale format, with five choices for each item ranging from strongly disagree to strongly agree, represented by numbers one to five (Table 7.7). The Likert scale format was used in the MFQ so as to allow room for manoeuvre, while at the same time maintaining control over the possible responses (Bloomer, 2010).

Table 7.7*Sample MFQ Question with Five-Level Likert Scale Format*

	1	2	3	4	5
Me encantan las clases de francés en el instituto.					

The MFQ was based primarily on Ryan's (2009) study, which replicated and expanded on the work of Dörnyei and his associates in Hungary (Dörnyei & Csizér, 2002; Csizér & Dörnyei, 2005a, 2005b; Dörnyei et al., 2006), a series of large-scale studies which used the L2 ideal selves theory as a framework in an attempt to validate the L2MSS theory. Ryan's study was chosen as it has been replicated and used extensively by numerous researchers to investigate the L2MSS in a number of different contexts including Japan, China and Iran (Ryan, 2008; Ryan, 2009; Taguchi et al., 2009), Pakistan (Islam et al., 2013), Spain (Brady, 2015) and Saudi Arabia (Moskovsky et al., 2016).

The questionnaire included a total of nine categories, which were made up of multi-item scales (Dörnyei & Csizér, 2012) of between four to fourteen items:

Ideal L2 Self (5)

The "Ought to" Self (7)

Language Anxiety (5)

Interest in Foreign Languages (6)

L2 Self Confidence (4)

Instrumentality: Prevention (5)

Instrumentality: Promotion (6)

Attitude towards Learning (8-14)²

Intended Learning Effort (5)

The Spanish version of the questionnaire was adopted from Brady (2015), which had undergone three important stages to ensure its validity. Firstly, it had been translated by the author and subsequently revised by two Spanish translators and three Spanish colleagues to check for potential errors. Secondly, it had been back-translated by a native English speaker to ensure equivalence of the two versions, as is advised by Dörnyei and

² The category "Attitude towards learning" varied to a greater degree, given that it incorporates students' attitudes towards learning in their different CLIL subjects which were different across the three grades. See below for a full explanation.

Csizér (2012). Finally, it had been piloted in order to check the participant’s interpretation of the translated items.

It should be noted that full piloting of the questionnaire for this study was not possible for two main reasons. Firstly, as mentioned above, the participants in this study were the only ones in the region who had the necessary profile required, namely studying content subjects through both English and French in secondary education. Furthermore, the dates which were initially offered by the schools to carry out the data collection did not allow for enough time to carry out a pilot test with a similar group of participants before the main data collection took place. As a result, the best piloting scenario available to the researcher was that of a think-aloud protocol (Dörnyei & Csizér, 2012). This involved having three colleagues answer the items in the questionnaire and provide feedback, after which the questionnaire was further revised prior to administration.

For the current study, Brady’s (2015) translated MFQ was adapted in two main ways. Firstly, given that the participants in the original study were in tertiary level education, three questions in the category “Attitude to Learning: Past and Present” referred to their past experiences of learning English in secondary school. In the current study, these questions were rewritten in the present tense given that the participants in this study were still in secondary school, as shown in Table 7.8.

Table 7.8

Changes to the Category “Attitude to Learning: Past and Present”

Brady (2015)	Present study
Me encantaban las clases de inglés en el instituto. [I loved English classes at school]	Me encantan las clases de inglés/francés en el instituto. [I love English/French classes at school]
He tenido profesores muy buenos de inglés. [I’ve had very good English teachers]	Tengo profesores muy buenos de inglés/francés. [I have very good English/French teachers]
Aprendí mucho inglés en el instituto. [I learned a lot of English at school]	Aprendo mucho inglés/francés en el instituto. [I learn a lot of English/French at school]

Secondly, additional questions were also added to this category (six to eight in the English MFQ and two in the French MFQ) in order to tap into the participants’ interest

in their CLIL subjects: Physics and Chemistry, Economics, Physical Education, and Technology (English) and Geography and History (French), as shown in Table 7.9. In addition to the original questions “Learning English/French is very interesting” and “I enjoy English/French lessons”, English/French was replaced with the relevant CLIL subjects.

Table 7.9

Additional Questions in the Category “Attitude to Learning: Past and Present”

English MFQ	French MFQ
Aprender <u>inglés</u> es muy interesante. [Learning English is very interesting]	Aprender <u>francés</u> es muy interesante. [Learning French is very interesting]
+ Aprender <u>física y química</u> / <u>economía</u> / <u>educación física</u> / <u>tecnología en inglés</u> es muy interesante. [Learning <u>Physics and Chemistry</u> / <u>Economics</u> / <u>Physical Education</u> / <u>Technology</u> in English is very interesting]	+ Aprender <u>geografía e historia en</u> <u>francés</u> es muy interesante. [Learning Geography and History in French is very interesting]
Yo disfruto en una clase de <u>inglés</u> . [I enjoy English lessons]	Yo disfruto en una clase de <u>francés</u> . [I enjoy French lessons]
+ Yo disfruto en una clase de <u>física y</u> <u>química</u> / <u>economía</u> / <u>tecnología</u> . [I enjoy <u>Physics and Chemistry</u> / <u>Economics</u> / <u>Technology</u> lessons]	+ Yo disfruto en una clase de <u>geografía e historia</u> . [I enjoy Geography and History lessons]

In addition to these extra questions which were specific to the students’ CLIL classes in each language, there were three other differences between the English and French questionnaires. Firstly, in part 1, question 7 dealt with the participants’ final grade in their language and CLIL subjects. Evidently, the English questionnaire asked about classes taught through English whereas the French questionnaire asked about classes taught through French. Secondly, throughout the MFQ the word “English” was replaced with “French” for the French version:

EN: Estudiar inglés es una pérdida de tiempo.

[Studying English is a waste of time]

FR: Estudiar francés es una pérdida de tiempo.

[Studying French is a waste of time]

Finally, the order of the questions in part 2 of the questionnaires were randomized, and therefore given in a different order in the MFQ of each language.

7.3.2. Procedure

As mentioned above, the main criterion for taking part in the study was that the participants must be enrolled in 9th, 10th or 11th grade in the selected schools, and therefore be studying content classes through both English and French. In other words, convenience sampling was used in this study (Dörnyei, 2007), as the students who participated all possessed the key characteristic relevant to the study: taking CLIL classes in English and French (Aiken, 1997). As is pointed out by Hatch and Lazaraton (1991), it was also necessary for statistical consideration to be taken into account, with the sample including more than 30 people. This condition was met both in the overall number of participants ($N = 91$) as well as the subgroups in 9th grade ($n = 42$), 10th grade ($n = 41$), 11th grade ($n = 40$), and students who participated in both 10th and 11th grade ($n = 32$).

Prior to the study, a consent form was signed by the directors of each school (Appendix G), which outlined the purpose of the study, as well as details regarding the data collection, confidentiality and results of the study.

The participants took the tests during their normal class time: the English-focused tests during their English-language class (approximately 50 minutes) and the French-focused tests during their French-language class (approximately 50 minutes).

Before taking the tests, all participants signed a consent form (Appendix H), indicating that they gave their approval for their data to be used in the study. They were informed that steps would be taken to ensure their confidentiality, including keeping the data in a safe location, not allowing anyone else to access the data, and not using the names of individuals in reporting the findings (Macksoud, 2010). The students were also informed that the results would be fully confidential, and that their personal data would not be used or distributed to other parties. As is pointed out by Dörnyei and Taguchi (2009), when it is necessary to link data from several tests to one participant, and even more so with longitudinal data, it is not possible for the tests to be completely anonymous.

The participants were thus instructed to include their names on tests, and it was explained that the researcher would then replace their names with a participant number. They were also reminded that the tests were not related to their school classes or results, and that their teachers would not have access to their tests.

Participants then completed the three tests in the following order for each language: the C-test (approximately 20 minutes), the LAT (approximately 15 minutes) and the questionnaire (approximately 10 minutes). For all tasks, precise instructions were given both orally and in written form in the students' L1 to ensure that they understood what they had to do. For the LATs, following the procedure detailed in Canga Alonso (2017), the time was controlled by the researcher and the participants were given two minutes to write as many words as possible for each prompt in the abovementioned word association task.

7.4. Data Analysis

The following sections outline the procedures that were used for the data analysis of the C-tests, LATs and questionnaires.

7.4.1. C-tests

Following the procedure outlined in Daller et al. (2003), in the marking of the C-tests, only exact solutions as those found in the source text were accepted. Each participant received one point for each correct word, for a score out of 80 for each C-test. These results were used to determine the participants' proficiency level. It should be noted that, given that the French and English C-tests have been taken from different sources, the participant's scores in each C-test are not comparable, that is, a higher or lower score in one C-test does not necessarily mean a higher or lower level of English or French. The results were used as an independent proficiency measure for each language, with the results of the English C-test being analysed alongside the English LAT and questionnaire, and the French C-test being analysed alongside the French LAT and questionnaire. This was done in the analysis of RQ2.2, as detailed below, in order to determine the relationship between language level, LA and language learning motivation, and in RQ3.1 and RQ4.1, in order to determine any differences in language level for the groups under analysis which may have impacted the results of the LAT.

7.4.2. Lexical Availability Tasks

In order to calculate the totals, means, standard deviations, and maximum and minimum values of LA for each language, learners' lemmatised word responses were

typed into Excel files, as detailed in Jiménez Catalán and Fernández Fontecha (2019). Criteria applied in the lemmatisation were adopted from Jiménez Catalán and Agustín Llach (2017) and included the following:

- i. Correction of spelling errors
- ii. Deletion of unintelligible words and Spanish L1 words, English words from the French test and French words from the English test
- iii. Repetition of the same word in the same prompt being counted only once
- iv. Lexical phrases as well as compound words being lemmatized as one lexical unit and counted as one word (e.g., orange-juice)
- v. Deletion of proper nouns (e.g., Greta Thumberg)
- vi. Changing plural words into the singular form, except in cases where the word is always plural (e.g., potatoes to potato)

To have a database of the excluded words, all removed items were also classified according to their exclusion using the following five categories: Spanish L1 words, French words (in the English LAT)/English words (in the French LAT), proper nouns, repeated words and other. Table 7.10 outlines each category alongside an example.

Table 7.10

Categorisation of Excluded Words in the LAT

From Spanish	toro [bull]	Proper Nouns	Fanta
From French	jambon [ham]	Repetition	basketball (included twice)
From English	sausage	Other	fuieds (unintelligible)

Following the preparation of the files, the data from the LAT were analysed using WordSmith Tools (Version 5), VocabProfile and SPSS (Version 26). For each prompt, words were counted and grouped into alphabetical and frequency lists using WordSmith Tools. This was done both with all participants and with each of the subgroups.

In order to answer RQ1.1, which asked if there were quantitative differences in the words retrieved by participants in English as compared to French at each testing period, paired samples t-tests, in the case of normally distributed data, and Wilcoxon signed-rank tests, in the case of non-normally distributed data, were carried out to compare the results of each LAT at each grade, given that two sets of scores obtained from the same group were being compared (Dörnyei, 2007). To address cross-sectional

and longitudinal differences in each language, independent samples t-tests, in the case of normally distributed data, and Mann-Whitney U tests, in the case of non-normally distributed data, were carried out to compare the cross-sectional data from 9th and 10th grade students, given that the scores of different groups were being compared, while paired samples t-tests, in the case of normally distributed data, and Wilcoxon signed-rank tests, in the case of non-normally distributed data, were carried out to compare the longitudinal data from 10th and 11th grade students, given that two sets of scores obtained from the same group were being compared (Dörnyei, 2007). The results of the above tests were also used to address RQ4.3, which asked if there was a quantitative difference in the number of words produced by students in English and French when the prompt is related to a CLIL class taken in either English or French at each testing period. Paired samples t-tests, given that the data were all normally distributed, were carried out on the three out of five prompts which were selected in order to tap into the content related vocabulary of the students CLIL classes: *Sports & Physical Activities* (Physical Education in English), *Environment & Climate* (Geography and History in French) and *Economy & Money* (Economics in English). For each grade, results in English were compared with those in French, in order to determine if there statistically significant differences between the languages in these specific prompts. Longitudinal differences in RQ4.3 were also investigated using paired-samples t-tests, comparing 10th and 11th grade participants' performance first in English and then in French.

Quantitative differences in the LAT by male and female students were addressed both across languages in RQ3.1.1 and within languages in RQ3.1.2. Regarding RQ3.1.1, which asked if there were quantitative differences between the male and female students in each grade across languages in the English and French LATs, paired samples t-tests, in the case of normally distributed data, and Wilcoxon signed-rank tests, in the case of non-normally distributed data, were carried out to compare the overall number of responses and the number of responses for each prompt produced by male participants in English and French and female participants in English and French in each of the three grades. To analyse the cross-sectional and longitudinal data, two-way mixed ANOVAS were carried out comparing participants in 9th and 10th grade, including one within-subjects factor (language) and one between-subjects factor (time), and two-way repeated ANOVAs were carried out comparing participants in 10th and 11th grade, with two within-subjects factors (time and language). This was done first for males and then for females in each case. Regarding RQ3.1.2, which asked if there were quantitative differences

between the male and female students in each grade within languages in the English and French LATs, independent samples t-tests, in the case of normally distributed data, and Mann-Whitney U tests, in the case of non-normally distributed data, were carried out. This was again carried out both for the whole LAT as well as for each prompt, as well as for students in each grade. To analyse the cross-sectional and longitudinal data, independent samples t-tests / Mann-Whitney U tests, were carried out comparing male and female participants in 9th and 10th grade, while two-way mixed ANOVAs, including one within-subjects factor (gender) and one between-subjects factor (time), were carried out comparing male and female participants in 10th and 11th grade, first in English and then in French in each case.

In order to answer RQ1.2, which asked if there were qualitative differences in the words produced in each language at each testing period, Wordsmith Tools was used to determine the most frequent first word for each prompt and the ranking of the most and least productive prompts and VocabProfile was used to analyse lexical sophistication, based on the non-shared words of participants as well as the number of infrequent words in the production of each prompt. For English, word responses were classified according to frequency on the basis of the BNC-COCA corpora, while for French the French v.5 corpora was used. In both cases, words were classified into 26 bands: the 1000 most frequent words (K1), the second 1000 most frequent words (K2), the third 1000 most frequent words in (K3), and so forth; and off-list words. It should be stressed that, given that the lists for each language come from different corpora, results in English and French are in this case not directly comparable. Instead, results are analysed independently for each language. The analysis was again done for students in each grade and compared cross-sectionally (9th to 10th grade) and longitudinally (10th to 11th grade).

The same procedure was used to answer RQ3.2, which asked if there were qualitative differences between male and female students in the words produced in each language at each testing period; RQ4.2., which asked whether there were qualitative differences in the words produced in English by students taking different CLIL classes at each testing period; and RQ4.4, which asked if there were qualitative difference in the words produced by students in English and French when the prompt was related to a CLIL class taken in either English or French at each testing period.

7.4.3. Questionnaires

The data collected by the administration of the questionnaire were analysed using SPSS. Numerical values were assigned to the 5 choices on the Likert scale: numerical value 1 was assigned to strongly agree, 2 to agree, 3 to somewhat agree/somewhat disagree, 4 to disagree, and 5 to strongly disagree. Prior to analysis, data cleaning and manipulation were carried out, and negatively worded items were re-coded by being reversed before the analysis (Dörnyei, 2003). As is typical in SLA, the cut-off point for concluding that a result was statistical was for the p-value to be below $\alpha = .05$ (Larson-Hall, 2012).

In order to answer RQ2.1, which asked if there was a significant difference between the participants' language learning motivation towards English as compared to French, Wilcoxon signed-rank tests were conducted for students in each grade, given that the data at hand was ordinal so consequently could not be tested using parametric tests (Field, 2009). Pearson Correlations, in the case of normally distributed data, and Spearman correlations, in the case of non-normally distributed data or ordinal data, were then used to address RQ2.2, which asked if there was a relationship between the results of the participants' LA task, language level and their responses in the MFQ, first for English and then for French and for each grade, and in RQ4.1., which asked if there were quantitative differences in the language level, LA and motivation of students taking different CLIL classes at each testing period. Effect sizes for R^2 were defined by Cohen (1992, as cited by Larson-Hall, 2015) as $R^2 = .01$ (small effect), $R^2 = .09$ (medium effect), and $R^2 = .25$ (large effect). Thus, the values for Pearson's r and Spearman's ρ were interpreted as followed: below .29 = small correlation, between 0.30 and 0.49 = moderate correlation, and between 0.50 and 1 = high correlation (Larson-Hall, 2012).

With regard to RQ3.3, quantitative differences in the language learning motivation by male and female students were addressed both across languages in RQ3.3.1 and within languages in RQ3.3.2. Regarding RQ3.3.1, which asked if there were quantitative differences between the male and female students in each grade across languages in English and French with regards to language learning motivation, Wilcoxon signed-rank tests were carried out, given that two sets of scores obtained from the same group were being compared (Dörnyei, 2007), for example, English language learning motivation versus French language learning motivation in 9th grade male students. Regarding RQ3.3.2 which asked if there were quantitative differences between the male and female students in each grade within languages in English and French with regards to language

learning motivation, Mann-Whiney U tests were carried out, given that the results of two independent groups were being compared (Dörnyei, 2007), for example, 9th grade male students versus 9th grade female students in English. For both RQ3.3.1 and RQ3.3.2, participants were also compared cross-sectionally (9th to 10th grade) by means of Mann-Whiney U tests and longitudinally (10th to 11th grade) by means of Wilcoxon signed-rank tests. In each case, the analyses focused on male students in English, male students in French, female students in English and female students in French. To address RQ3.3.1, a comparison was then carried out on the results of male students in English versus French, and of female students in English versus French. To address RQ3.3.2 a comparison was carried out on the results of male versus female students in English, and male versus female students in French.

Finally, regarding RQ4.1, which investigated quantitative difference in the motivation of students taking different CLIL classes at each testing period, non-parametric Mann-Whitney U tests were carried out to determine if differences between the science and economics students in terms of their English language learning motivation were statistically significant, while Kruskal-Wallis H tests were used to determine if there were longitudinal differences from 10th to 11th grade.

Chapter 8: Results

This section presents the results of the data analysis, addressing each area and research question in turn: Target Language Influence on LA, Target Language Influence on Language Learning Motivation, Gender, and CLIL.

8.1. Target Language Influence on Lexical Availability

The first research question aimed to determine whether there were quantitative (RQ1.1) and qualitative (RQ1.2) differences in the participants' LA in English as compared to French in each grade at each testing period.

8.1.1. *Quantitative Differences in English and French LA*

Research question 1.1 asked whether there were quantitative differences in the words retrieved by participants in English as compared to French in each grade at each testing period. The descriptive statistics for the overall LAT for each language, as shown in Table 8.1, indicated that participants produced a higher number of words in English than they did in French, both in the overall LAT and in each of the individual prompts in all grades.

Following the analysis of the descriptive statistics, the normality of each prompt and the overall LAT in each language was assessed for each grade (Table 8.2). As shown, the data were normally distributed in all cases except for the prompt *Animals* in French in 9th grade and in English in 10th grade. Thus, paired samples t-tests, in the case of normally distributed data, and Wilcoxon signed-rank tests, in the case of the prompt *Animals* in 9th and 10th grade which was non-normally distributed, were carried out using SPSS to compare the number of words produced in each language and determine whether the differences observed were statistically significant. This was done both for the whole LAT as well as for each individual prompt, and for each grade under analysis.

Results revealed that the differences were statistically significant in all cases, with participants producing a statistically significant higher number of tokens in English than in French in all five prompts and overall (Table 8.3). The results are presented visually below in Figure 8.1.

Table 8.1*Descriptive Statistics for Lexical Availability in English and French³*

9 th grade								
English (<i>n</i> = 39)					French (<i>n</i> = 42)			
Min	Max	<i>M</i>	<i>SD</i>	Prompt	Min	Max	<i>M</i>	<i>SD</i>
1	32	19.08	7.10	Animals	2	29	10.33	5.62
0	35	19.77	8.61	Food & Drink	1	25	9.83	4.93
3	29	14.62	5.90	Sport & PA	0	18	9.49	4.05
4	34	17.31	7.63	Environment & Climate	1	23	10.38	5.20
0	24	10.44	6.01	Economy & Money	0	18	7.86	5.00
21	151	81.20	31.23	Mean LAT	10	109	47.42	19.68
10 th grade								
English (<i>n</i> = 41)					French (<i>n</i> = 41)			
Min	Max	<i>M</i>	<i>SD</i>	Prompt	Min	Max	<i>M</i>	<i>SD</i>
10	28	18.22	4.83	Animals	2	19	8.54	3.47
12	35	21.61	5.29	Food & Drink	2	23	10.80	5.28
9	27	16.61	4.56	Sport & PA	6	17	11.22	2.56
2	31	18.61	5.86	Environment & Climate	3	23	10.41	4.75
4	26	13.95	4.79	Economy & Money	2	17	10.41	3.74
46	146	89.00	20.04	Mean LAT	25	83	51.39	15.28
11 th grade								
English (<i>n</i> = 40)					French (<i>n</i> = 36)			
Min	Max	<i>M</i>	<i>SD</i>	Prompt	Min	Max	<i>M</i>	<i>SD</i>
7	30	19.97	5.65	Animals	4	18	11.06	3.84
8	36	22.75	7.22	Food & Drink	1	23	11.28	5.63
9	27	17.35	4.65	Sport & PA	4	15	10.17	3.12
4	32	18.40	7.40	Environment & Climate	2	21	12.08	5.99
2	24	13.83	4.92	Economy & Money	0	20	9.47	4.52
48	142	92.30	24.10	Mean LAT	16	86	54.05	18.84

³ Types results are not considered here given that individual results and not aggregated results are provided. As a result, given the nature and requirements of the task, types and tokens coincide in each individual set of responses per each prompt.

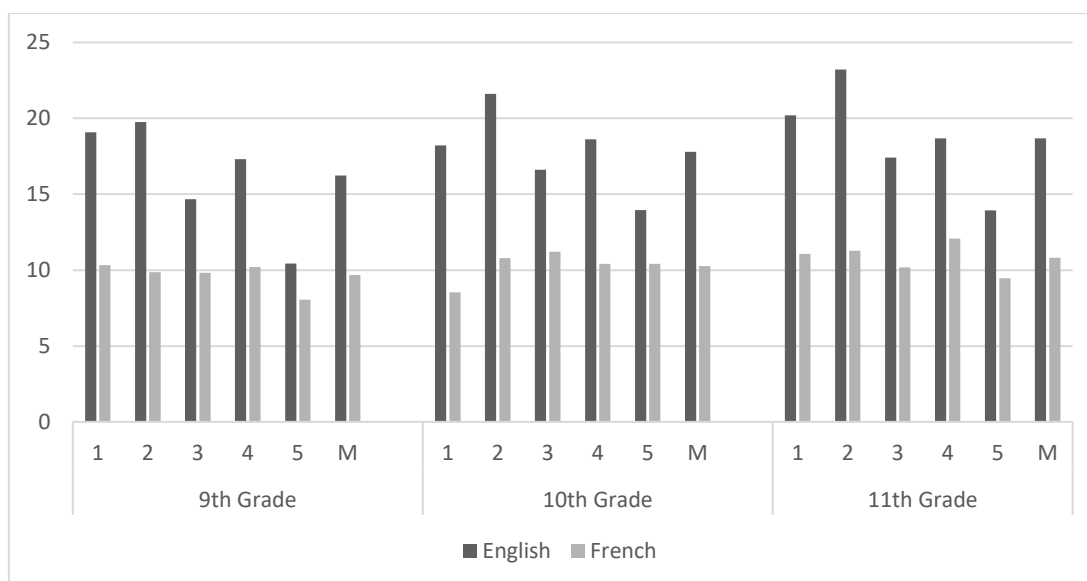
Table 8.2*Normality Tests for Lexical Availability in English and French*

9 th grade						
English				French		
<i>W</i>	<i>df</i>	<i>p</i>	Prompt	<i>W</i>	<i>df</i>	<i>p</i>
.970	39	.377	Animals	.889	42	.001
.975	39	.541	Food & Drink	.952	41	.079
.963	39	.224	Sport & Physical Activities	.980	41	.686
.979	39	.681	Environment & Climate	.979	42	.607
.977	39	.591	Economy & Money	.952	42	.078
.977	39	.583	LAT Mean	.957	42	.119
10 th grade						
English				French		
<i>W</i>	<i>df</i>	<i>p</i>	Prompt	<i>W</i>	<i>df</i>	<i>p</i>
.944	41	.043	Animals	.962	41	.183
.972	41	.410	Food & Drink	.972	41	.399
.963	41	.206	Sport & Physical Activities	.976	41	.518
.978	41	.604	Environment & Climate	.954	41	.093
.978	41	.593	Economy & Money	.977	41	.552
.982	41	.747	LAT Mean	.974	41	.477
11 th grade						
English				French		
<i>W</i>	<i>df</i>	<i>p</i>	Prompt	<i>W</i>	<i>df</i>	<i>p</i>
.969	40	.345	Animals	.949	36	.096
.979	40	.636	Food & Drink	.952	36	.124
.968	40	.305	Sport & Physical Activities	.954	36	.139
.950	40	.074	Environment & Climate	.941	36	.056
.975	40	.496	Economy & Money	.978	36	.692
.972	40	.422	LAT Mean	.956	36	.165

Table 8.3*Differences in Lexical Availability in English and French*

9 th grade							
Prompt	<i>M</i>		<i>SD</i>		<i>t / z</i>	<i>df</i>	<i>p</i>
	English	French	English	French			
Animals*	19.08	10.33	7.10	5.62	-4.91		<.001
Food & Drink	19.76	9.87	8.73	5.11	8.06	37	<.001
Sport & Physical Activities	14.68	9.82	5.96	4.02	5.52	37	<.001
Environment & Climate	17.31	10.21	7.63	5.28	6.72	38	<.001
Economy & Money	10.44	8.05	6.01	4.84	3.44	38	.001
Mean LAT	16.24	9.67	6.24	4.02	8.69	38	<.001
10 th grade							
Prompt	<i>M</i>		<i>SD</i>		<i>t / z</i>	<i>df</i>	<i>p</i>
	English	French	English	French			
Animals*	18.22	8.54	4.83	3.47	-5.57		<.001
Food & Drink	21.61	10.80	5.29	5.28	12.93	40	<.001
Sport & Physical Activities	16.61	11.22	4.56	2.56	7.74	40	<.001
Environment & Climate	18.61	10.41	5.86	4.75	10.67	40	<.001
Economy & Money	13.95	10.41	4.79	3.74	4.51	40	<.001
Mean LAT	17.80	10.27	4.00	3.05	15.16	40	<.001
11 th grade							
Prompt	<i>M</i>		<i>SD</i>		<i>t</i>	<i>df</i>	<i>p</i>
	English	French	English	French			
Animals	20.19	11.06	5.85	3.84	11.27	35	<.001
Food & Drink	23.22	11.28	6.88	5.63	14.19	35	<.001
Sport & Physical Activities	17.42	10.17	4.77	3.12	9.46	35	<.001
Environment & Climate	18.69	12.08	7.47	5.99	7.60	35	<.001
Economy & Money	13.94	9.47	5.11	4.52	5.56	35	<.001
Mean LAT	18.69	10.81	4.90	3.76	14.20	35	<.001

Note. * = non-parametric test used, given that the data were not normally distributed.

Figure 8.1*Lexical Availability in English and French*

Note. 1 = Animals, 2 = Food & Drink, 3 = Sport and Physical Activities, 4 = Environment & Climate, 5 = Economy & Money, M = Mean LAT

In order to determine if these results were consistent with the participants' self-reported language levels, Wilcoxon signed-rank tests were carried out to compare the participants' indicated language level in English and French, given that the data at hand was ordinal. Self-reported language levels were used here given that it was not possible to directly compare the results of one C-test and the other, as previous indicated. Results revealed statistically significant differences between the two languages in each grade, with students reporting a higher level in English ($Mdn = 3.00$, $SD = 1.08$ in 9th grade; $Mdn = 2.00$, $SD = .51$ in 10th grade; $Mdn = 3.00$, $SD = 1.28$ in 11th grade) than French ($Mdn = 3.00$, $SD = 1.19$ in 9th grade; $Mdn = 2.00$, $SD: .45$ in 10th grade; $Mdn = 3.00$, $SD = 1.46$ in 11th grade) in 9th ($z = -2.80$, $p = .005$), 10th ($z = -2.98$, $p = .003$) and 11th grade ($z = -2.21$, $p = .027$). This suggests that the students believe that their language level in English is higher than in French, which is consistent with their results on token production in the LAT. In addition, it is what would be expected given the higher number of hours of exposure to English as compared to French.

In order to determine whether there were cross-sectional (9th and 10th grade) and longitudinal differences (10th and 11th grade) in the English and French LATs at each testing period, independent samples t-tests / Mann-Whitney U tests and paired samples t-tests / Wilcoxon signed-rank tests, respectively, were carried out. As noted above, the

non-parametric Mann-Whitney U tests and Wilcoxon signed-rank tests were used in the case of the prompt *Animals* in French in 9th grade and the prompt *Animals* in English in 10th grade, given that the data were not normally distributed.

Firstly, as shown in Table 8.4 below, results of the independent samples t-test and Mann-Whitney U tests comparing participants in 9th grade with participants in 10th grade revealed that there was a statistically significant difference only in one of the five prompts in English: *Economy and Money*.

Table 8.4

Cross-Sectional Differences in the English and French LATs

9 th → 10 th grade							
English							
Prompt	M		SD		t / z	df	p
	9 th	10 th	9 th	10 th			
Animals*	19.08	18.22	7.10	4.83	-.695		.487
Food & Drink	19.77	21.61	8.61	5.29	-1.14	62.54	.257
Sport & Physical Activities	14.62	16.61	5.90	4.56	-1.69	78	.094
Environment & Climate	17.31	18.61	7.63	5.86	-.85	78	.393
Economy & Money	10.44	13.95	6.01	4.79	-2.89	78	.005
Mean LAT English	16.24	17.80	6.24	4.00	-1.32	64.23	.191
9 th → 10 th grade							
French							
Prompt	M		SD		t / z	df	p
	9 th	10 th	9 th	10 th			
Animals*	10.33	8.54	5.62	3.47	-.997		.319
Food & Drink	9.83	10.80	4.93	5.28	-.86	80	.390
Sport & Physical Activities	9.49	11.22	4.05	2.56	-2.31	67.62	0.24
Environment & Climate	10.38	10.41	5.20	4.75	-.03	81	.976
Economy & Money	7.86	10.41	5.00	3.74	-2.64	75.88	.010
Mean LAT French	9.59	10.27	3.92	3.05	-.88	81	.379

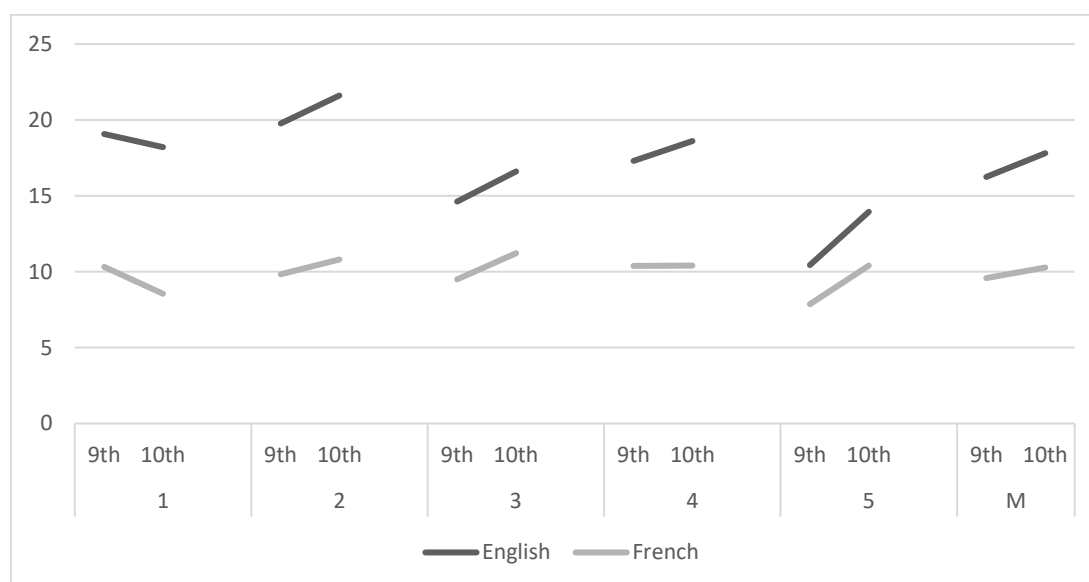
Note. * = non-parametric test used, given that the data were not normally distributed.

This finding is extremely important given that this prompt is a specific, content-relevant prompt linked to economics which was studied for the first time by students in 10th grade. In other words, while all prompt-relevant CLIL subjects showed no differences between the two groups, by studying economics through English the 10th grade, students had

acquired enough vocabulary to surpass the 9th grade students in this lexical domain. Regarding French, statistically significant differences were found in two prompts: *Sport & Physical Activities* and *Economy & Money*. These prompts are those which were specifically included to tap into content-related vocabulary of physical education and economics, which were studied not in French but in English. It is, however, possible that, given the large number of English-French cognates in these lexical domains, the students were able to rely on vocabulary acquired through their English-taught subjects in order to access more words in French. This will be discussed in greater detail in the qualitative analysis (Section 8.1.2). Figure 8.2. offers a visually representation of these results.

Figure 8.2

Cross-Sectional Differences in the English and French LATs



Note. 1 = *Animals*, 2 = *Food & Drink*, 3 = *Sport and Physical Activities*, 4 = *Environment & Climate*, 5 = *Economy & Money*, M = Mean LAT

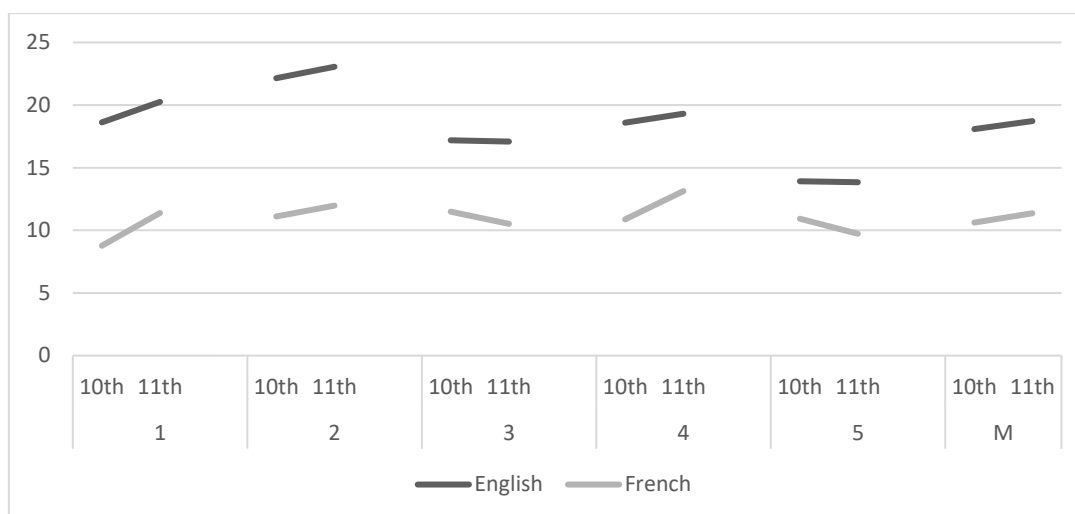
Secondly, as shown in Table 8.5 below, results of the paired samples t-tests and Wilcoxon signed-rank tests comparing the same participants in 10th and 11th grade showed that there were no statistically significant differences in any of the five prompts or for the overall English LAT. However, with regards to French, statistically significant differences were found in two of the five prompts, *Animals* and *Environment & Climate*, as well as in the overall mean for LA.

Table 8.5*Longitudinal Differences in the English and French LATs*

10 th → 11 th grade							
English (<i>n</i> = 32)							
Prompt	<i>M</i>		<i>SD</i>		<i>t</i> / <i>z</i>	<i>df</i>	<i>p</i>
	10 th	11 th	10 th	11 th			
Animals*	18.63	20.25	4.84	5.81	-1.93		.053
Food & Drink	22.16	23.06	5.45	7.38	-.954	31	.347
Sport & Physical Activities	17.19	17.09	4.74	4.48	.152	31	.880
Environment & Climate	18.59	19.31	6.34	7.69	-.833	31	.411
Economy & Money	13.91	13.84	5.07	5.08	.098	31	.923
Mean LAT English	18.09	18.71	4.28	5.01	-1.37	31	.178
10 th → 11 th grade							
French (<i>n</i> = 30)							
Prompt	<i>M</i>		<i>SD</i>		<i>t</i>	<i>df</i>	<i>p</i>
	10 th	11 th	10 th	11 th			
Animals	8.77	11.40	3.42	3.86	-4.32	29	<.001
Food & Drink	11.10	11.97	5.79	5.78	-1.31	29	.197
Sport & Physical Activities	11.50	10.53	2.17	3.08	1.99	29	.056
Environment & Climate	10.87	13.13	4.47	5.77	-3.36	29	.002
Economy & Money	10.93	9.73	3.72	4.77	1.47	29	.152
Mean LAT French	10.63	11.35	3.06	3.78	-2.06	29	.048

Note. * = non-parametric test used, given that the data were not normally distributed.

Notably, the prompt *Environment & Climate* was included to tap into vocabulary related to geography, which was studied through French in 10th grade but not in 11th grade. Given that the first data collection took place in January while the participants were in 10th grade, it is possible that they continued to acquire content-relevant vocabulary in this subject over the rest of this academic year (around five months of instruction) and retained this vocabulary up until the second data collection the following year. It should also be noted that while a statistically significant difference was found in the overall LAT for French, this was not the case for English. These findings are presented visually in Figure 8.3.

Figure 8.3*Longitudinal Differences in the English and French LATs*

Note. 1 = Animals, 2 = Food & Drink, 3 = Sport and Physical Activities, 4 = Environment & Climate, 5 = Economy & Money, M = Mean LAT

In order to determine if these findings were related to the students' language level, the normality of the results of the C-tests in each language were first assessed for each grade (Table 8.6). C-tests were used here, given that the English C-test could be compared from one grade to the next and the French C-test from one grade to the next, rather than comparing the English and French C-tests. As shown below, data were normally distributed at all grades. Thus, parametric tests were used in analyses using the C-tests, unless other non-normally distributed data were being compared.

Table 8.6*Normality Tests for C-tests*

9 th Grade					
English C-test			French C-test		
<i>W</i>	<i>df</i>	<i>p</i>	<i>W</i>	<i>Df</i>	<i>p</i>
.960	41	.163	.972	42	.386
10 th Grade					
English C-test			French C-test		
<i>W</i>	<i>df</i>	<i>p</i>	<i>df</i>	<i>W</i>	<i>p</i>
.978	41	.978	41	.978	41
11 th Grade					
English C-test			French C-test		
<i>W</i>	<i>df</i>	<i>p</i>	<i>df</i>	<i>W</i>	<i>p</i>
.970	40	.970	40	.970	40

Regarding cross-sectional differences, results of the paired samples t-tests showed that there were no statistically significant differences between the participants' performance on the C-test in 9th grade ($M = 29.15$, $SD = 12.38$) and 10th grade ($M = 34.05$, $SD = 11.87$) in English ($t(80) = -1.83$, $p = .071$) and in 9th grade ($M = 30.43$, $SD = 11.12$) and 10th grade ($M = 35.44$, $SD = 12.57$) in French ($t(81) = -1.92$, $p = .058$). This is consistent with the results of the overall LAT, where there were no statistically significant differences in either language between 9th and 10th grade students. Thus, the statistically significant differences found in the individual prompts *Economy & Money* in English and *Animals* and *Environment & Climate* in French cannot be attributed to differences in language level. Regarding longitudinal differences, despite the different findings for each language on the LAT overall, results of the paired samples t-tests showed that there were statistically significant differences between the participants' performance on the C-test in 10th grade ($M = 33.78$, $SD = 12.84$) and 11th grade ($M = 36.91$, $SD = 12.61$) in English ($t(31) = -2.64$, $p = .013$) and in 10th grade ($M = 35.97$, $SD = 12.57$) and 11th grade ($M = 42.60$, $SD = 9.04$) in French ($t(29) = -2.19$, $p = .036$). Thus, although overall language level appeared to improve from one grade to the next in both languages, this did not entail an overall improvement in the LAT for both languages, where a statistically significant difference was found in the overall LAT for French but not for English. It does thus not seem to be the case that the improvement on the French LAT can simply be attributed to an improvement in language level. However, given the difference between the two languages in the LAT and the fact that retrieval in English was much higher at both data collections, as discussed above, it is possible that there was something of a ceiling effect at play in the English LAT. In other words, given that a particularly high number of words were retrieved in English in 10th grade, there was less room for improvement. On the other hand, as significantly fewer words were retrieved in French in 10th grade, there was a higher likelihood that the participants would improve their vocabulary over the course of the year.

In summary, participants produced a statistically significant higher number of tokens in English than in French in all five prompts and overall in 9th, 10th and 11th grade. This was consistent with their self-reported language levels, as students reported statistically significant higher language levels in English than French in all grades. Regarding the cross-sectional and longitudinal analyses, no statistically significant differences were found between 9th and 10th grade students in the overall LAT in either language or in the performance on the English and French C-tests from one grade to the

next. However, a statistically significant difference was found in just one of the five prompts in English, namely *Economy and Money*, with 10th grade students producing a higher number of tokens. This is very likely related to the fact that 10th grade students studied economics through English, whereas the 9th grade students did not. Statistically significant differences were also found in two prompts in the French LAT, *Sport & Physical Activities* and *Economy & Money*, which were notably prompts included to tap into content-related vocabulary of physical education and economics, studied not in French but in English. As for longitudinal differences, no statistically significant differences were found in the five prompts or the overall English LAT, but there was a statistically significant difference in the participants' performance on the English C-test. In French, however, statistically significant differences were found in two of the five prompts, *Animals* and *Environment & Climate*, as well as in the overall mean for LA, and in the French C-test.

8.1.2. Qualitative Differences in English and French LA

Research question 1.2 aimed to determine whether there were qualitative differences in the words retrieved by participants in English as compared to French at each testing period. In order to address this, a qualitative analysis was carried out to identify the frequency of first word responses for each prompt, the most and least productive prompts, and the lexical sophistication of each language based on the non-shared words of participants as well as the number of infrequent words in the production of each prompt.

Firstly, with regard to the most frequent first word for each prompt, Wordsmith Tools Version 5 was used to analyse the most frequent first word responses retrieved in English and French for each prompt, and the number and percentage of participants that retrieved the response (Table 8.7). Results showed that the first word was the same in English and French for all three grades in three out of the five prompts: *Animals* (“dog”/ “chien”), *Food & Drink* (“water” / “eau”) and *Environment & Climate* (“sun” / “soleil”). However, with regard to the prompts *Sport & Physical Activities* and *Economy & Money*, some differences were found. Regarding *Sport & Physical Activities*, while the most common first word in English and French for 9th and 11th grade was “football”, for 10th grade students it was “football” in English, but “basketball” in French. This finding will be discussed in greater detail in Section 8.3 with regards to gender. Regarding *Economy & Money*, the most frequent first word was different for each group: for 9th grade students,

it was “dollar” in English and “euro” in French, for 10th grade students it was “economy” in both languages, and for 11th grade students it was “money” in both languages. Notably, for this prompt participants in 9th grade chose a currency in each language (“dollar” / “euro”) while participants in 10th and 11th grade relied on vocabulary that was in the title of the prompt itself (“economy” / “money”).

Table 8.7*Most Frequent First Words for Each Prompt*

9 th grade				
	English		French	
Prompt	Lexical Unit	N of informants	Lexical Unit	N of informants
1	<i>Dog</i>	16 (41%)	<i>Chien</i>	16 (38.1%)
2	<i>Water</i>	15 (39.5%)	<i>Eau</i>	11 (26.8%)
3	<i>Football</i>	11 (28.2%)	<i>Football</i>	15 (37.5%)
4	<i>Sun</i>	12 (30.8%)	<i>Soleil</i>	23 (54.8%)
5	<i>Dollar</i>	7 (19.4%)	<i>Euro</i>	10 (24.4%)
10 th grade				
	English		French	
Prompt	Lexical Unit	N of informants	Lexical Unit	N of informants
1	<i>Dog</i>	23 (56.1%)	<i>Chien</i>	20 (48.8%)
2	<i>Water</i>	12 (29.3%)	<i>Eau</i>	17 (41.5%)
3	<i>Football</i>	14 (34.2%)	<i>Basket</i>	12 (29.3%)
4	<i>Sun</i>	10 (24.4%)	<i>Soleil</i>	13 (31.7%)
5	<i>Economy</i>	10 (24.4%)	<i>Économie</i>	11 (26.8%)
11 th grade				
	English		French	
Prompt	Lexical Unit	N of informants	Lexical Unit	N of informants
1	<i>Dog</i>	13 (32.5%)	<i>Chien</i>	13 (36.1%)
2	<i>Water</i>	12 (30%)	<i>Eau</i>	18 (50%)
3	<i>Football</i>	16 (40%)	<i>Football</i>	9 (25%)
4	<i>Sun</i>	9 (22.5%)	<i>Soleil</i>	14 (38.9%)
5	<i>Money</i>	11 (27.5%)	<i>Argent</i>	18 (51.4%)

Note. Percentages indicate the number of participants who produced the given word out of the total number of participants who produced words for the given prompt. Prompt 1 = *Animals*, Prompt 2 = *Food & Drink*, Prompt 3 = *Sport and Physical Activities*, Prompt 4 = *Environment & Climate*, Prompt 5 = *Economy & Money*.

Secondly, the prompts were analysed to determine the ranking of the most and least productive prompts for both English and French in each grade (Table 8.8). Results showed a great degree of variation in the productivity of each prompt in English as compared to French. One of the most striking findings is the differences in the productivity of the content-relevant prompts. For English, prompts related to subjects

studied through English (*Sport & Physical Activities* and *Economy & Money*) were the least productive at all three grades. However, for French, the content-relevant prompt (*Environment & Climate*) was found to be the most productive for 9th grade and 11th grade students. Thus, despite studying physical education and economics through English, vocabulary related to these domains did not surpass other more productive ones such as *Food & Drink* or *Animals* in the English LAT. On the other hand, studying geography and history through French potentially provided participants with enough lexical items to exceed the number produced for more productive prompts. This was not, however, the case in 10th grade.

Table 8.8*Ranking of Most and Least Productive Prompts*

9 th grade				
Rank	English	Tokens	French	Tokens
1	Food & Drink	771	Environment & Climate	436
2	Animals	744	Animals	434
3	Environment & Climate	675	Food & Drink	403
4	Sport & Physical Activities	570	Sport & Physical Activities	389
5	Economy & Money	407	Economy & Money	330
10 th grade				
Rank	English	Tokens	French	Tokens
1	Food & Drink	886	Sport & Physical Activities	460
2	Environment & Climate	763	Food & Drink	443
3	Animals	747	Environment & Climate	427
4	Sport & Physical Activities	681	Economy & Money	427
5	Economy & Money	572	Animals	350
11 th grade				
Rank	English	Tokens	French	Tokens
1	Food & Drink	910	Environment & Climate	435
2	Animals	799	Food & Drink	406
3	Environment & Climate	736	Animals	398
4	Sport & Physical Activities	694	Sport & Physical Activities	366
5	Economy & Money	553	Economy & Money	341

In order to investigate these differences further, an analysis was carried out on another factor which may have had an influence on the results: the presence of cognates. VocabProfile was used to determine the percentage of French cognates in the English LAT and the percentage of English cognates in the French LAT (Table 8.9). This feature works by analysing words which pertain to cognates in classified on-list items, which for English lie in the 1-11K bands and in French in the 1-5K bands.

Table 8.9*Cognates in the LAT*

	9 th grade		10 th grade		11 th grade	
	Prompt	Cognates	Prompt	Cognates	Prompt	Cognates
English	2	46%	2	44%	2	45%
	1	26%	4	21%	1	27%
	4	20%	1	25%	4	21%
	3	39%	3	38%	3	39%
	5	42%	5	49%	5	45%
French	4	34%	3	76%	4	27%
	1	26%	2	39%	2	41%
	2	39%	5	69%	1	15%
	3	83%	4	31%	3	78%
	5	75%	1	12%	5	69%

Note. Content-relevant prompts for each language and grade are marked in bold. Prompt 1 = *Animals*, Prompt 2 = *Food & Drink*, Prompt 3 = *Sport and Physical Activities*, Prompt 4 = *Environment & Climate*, Prompt 5 = *Economy & Money*.

As can be seen, in 9th grade, 46% of the items in prompt 2, *Food & Drink*, of the English LAT were French cognates, 34% of the items in prompt, *Environment and Climate*, of the French LAT were English cognates, and so on. Prompts are presented in order of the ranking of most to least productive as outlined above. As noted above, the content-relevant prompt *Environment & Climate* was found to be the most productive prompt in the French LAT for 9th and 11th, while the same was not true for 10th grade students. For these participants, the prompt ranked as the most productive for French, *Sport & Physical Activities*, is in fact that with the highest percentage cognates (76%) while the prompt that ranked as the least productive prompt, *Animals*, is the one with the lowest percentage of cognates (12%). This may indicate that the participants in this grade relied on English cognates in their lexical production during the French LAT, and could explain to some extent the variation in the productivity of each prompt in English as compared to French. While it is not yet possible to analyse the presence of Spanish cognates using VocabProfile, this could also be a potentially interesting factor in the LAT, given that it is very likely that the participants would have also relied on Spanish/English and Spanish/French cognates.

Finally, analyses were carried out to examine lexical sophistication based on the non-shared words of participants as well as the number of infrequent words in the production of each prompt. With regard to non-shared words, each prompt was analysed in turn to determine the number of words which were unique to one participant, and the

percentage of these non-shared words. For example, in the prompt *Economy & Money* in English, 9th grade students produced 114 non-shared words out of a total of 407 words, meaning that 28% of the words produced by the participants were non-shared words (Table 8.10).

Table 8.10*Non-shared Words*

9 th grade				
Rank	English	Non-shared Words	French	Non-shared words
1	Eco & Money	114/407 (28%)	Eco & Money	78/330 (23.6%)
2	Sport & PA	110/570 (19.2%)	Sport & PA	43/389 (11%)
3	Env & Climate	88/675 (13%)	Env & Climate	42/436 (9.6%)
4	Food & Drink	59/771 (7.6%)	Food & Drink	38/403 (9.4%)
5	Animals	51/744 (6.8%)	Animals	36/434 (8.2%)
10 th grade				
Rank	English	Non-shared words	French	Non-shared words
1	Eco & Money	139/572 (24.3%)	Eco & Money	72/427 (17%)
2	Sport & PA	111/681 (16.2%)	Env & Climate	51/427 (11.9%)
3	Env & Climate	108/763 (14.1%)	Food & Drink	44/443 (9.9%)
4	Animals	86/747 (11.5%)	Sport & PA	36/460 (7.8%)
5	Food & Drink	93/886 (10.4%)	Animals	26/350 (7.4%)
11 th grade				
Rank	English	Non-shared Words	French	Non-shared words
1	Eco & Money	151/553 (27.3%)	Eco & Money	69/341 (20.2%)
2	Env & Climate	105/736 (14.2%)	Env & Climate	66/435 (15.1%)
3	Sport & PA	82/694 (11.8%)	Food & Drink	54/406 (13.3%)
4	Food & Drink	76/910 (8.3%)	Sport & PA	31/366 (8.4%)
5	Animals	43/799 (5.3%)	Animals	22/398 (5.5%)

The prompts were then ranked to examine the prompts with the most and least non-shared words for each language. Three interesting observations were made with regard to the non-shared words. The first is the similarities which were found with regard to the ranking, with *Economy & Money* producing the highest number of non-shared words in both English (9th grade: 28%, 10th grade: 24.3%, 11th grade: 27.3%) and French (9th grade: 23.6%, 10th grade: 17%, 11th grade: 20.2%), and *Animals* producing the lowest number of non-shared words in almost all cases in both English (9th grade: 6.8%, 11th grade: 5.3%) and French (9th grade: 8.2%, 10th grade: 7.4%, 11th grade: 5.5%). This suggests that there may be less variety in semantic fields such as *Animals*, which generally has a high level of exposure, as compared with prompts such as *Economy & Money*, which in terms of

vocabulary may have received less exposure. Secondly, content-relevant prompts very often revealed a higher number of non-shared words across the three groups, with *Economy & Money*, *Environment & Climate* and *Sport & Physical Activities* often having the highest percentage of non-shared words. However, given that fact that this is the case for *Economy & Money* in 9th grade participants, who had not yet begun to study economics, and the fact that ranking was similar for both English and French, this is likely due to the actual content of the prompt rather than the fact the students studied prompt-relevant vocabulary in their CLIL classes. Finally, for almost all prompts, the participants produced a higher percentage of non-shared words in English than in French. Some minor exceptions to this were found in the prompts with fewer non-shared words such as *Animals* and *Food & Drink* in the case of 9th and 11th grade students and *Environment & Climate* in the case of 11th grade students, where participants produced a higher percentage of non-shared words in French than in English. The fact that there was generally a higher percentage of non-shared words in English indicates that, in addition to producing a statistically higher number of words in English than in French, participants also showed a higher level of lexical sophistication in English in terms of non-shared words.

The final part of the qualitative analysis consisted in an analysis of lexical sophistication in terms of the number of infrequent words produced in each language. This was done using VocabProfile, classifying learners' word responses according to frequency on the basis of the BNC-COCA corpora (English) and the French v.5 corpora (French) into 25 bands: the 1000 most frequent words in English/French (K1), the second 1000 most frequent words in English/French (K2), the third 1000 most frequent words in English/French (K3), and so forth; and off-list words. For ease of readability, the results below are presented across 6 categories: K1-K5, K6-K10, K11-K15, K16-K-20, K21-25 and off-list words.

It should be borne in mind that, given that the corpora for each language are evidently different, it is not possible to directly compare the following results between English and French. Thus, findings which show that participants produce a higher number of lower-frequency words in English or French does not mean that they have shown a higher level of lexical sophistication in one of the languages, as the corpora for each are inherently different. Instead, observations will be made first for English (Table 8.11), comparing the results between the five prompts to determine the number of infrequent words in each category. The same will then be done for French (Table 8.12).

Table 8.11*Frequency Distributions English*

9 th grade						
Prompt	K1-K5	K6-K10	K11-K15	K16-K20	K21-K25	Off-list
1	75.9	17.3	2.4	2.4	0.1	1.8
2	79.8	11.3	3.9	0.8	0	4
3	67.9	14.2	2	0	0	15.9
4	91.4	1.9	0.5	0	0	5.9
5	82.5	9.4	0	0	0	8.1
10 th grade						
Prompt	K1-K5	K6-K10	K11-K15	K16-K20	K21-K25	Off-list
1	77.9	16.3	3	0.7	0.1	1.8
2	77.7	12	3.3	0.7	0	6.2
3	67.7	10.7	1.4	0.3	0.3	19.5
4	87.7	3.9	0.4	0	0.2	7.7
5	84	5	0.3	0.2	0	10.4
11 th grade						
Prompt	K1-K5	K6-K10	K11-K15	K16-K20	K21-K25	Off-list
1	77.8	17	2	0.9	0	2.7
2	75	13.8	4.7	0.6	0.1	5.8
3	68	9.9	1.8	0.4	0	19.8
4	87.7	3.2	0.2	0	0.3	8.7
5	80.8	8.3	1.3	0	0	9.5

Note. Prompt 1 = *Animals*, Prompt 2 = *Food & Drink*, Prompt 3 = *Sport and Physical Activities*, Prompt 4 = *Environment & Climate*, Prompt 5 = *Economy & Money*.

Regarding English, it should firstly be noted that the less common semantic categories, *Environment & Climate* and *Economy & Money*, contained a higher percentage of words in the K-1 to K-5 lists than the other semantic categories. As shown, this band accounted for 91.4% of tokens in 9th grade, 87.7% of tokens in 10th grade and 87.7% of tokens 11th grade for *Environment & Climate*, and for 82.5% of tokens in 9th grade, 84% of tokens in 10th grade and 80.8% of tokens in 11th grade. This indicates that in addition to producing fewer words in these categories, the types of words that were produced were generally less sophisticated. Another interesting observation is the high percentage of off-list words in the category *Sport & Physical Activities*, making up 15.96%, 19.53% and 19.88% of the tokens in 9th, 10th and 11th grade respectively. In general, this can be attributed to the high number of multiword units within this category, which are analysed as off-list. For example, the token “iceskating” when broken up is composed of “ice”, a 1K token, and “skating”, a 5K token. However, when taken as a single token, it appears as an off-list word.

Table 8.12*Frequency Distributions French*

9 th grade						
Prompt	K1-K5	K6-K10	K11-K15	K16-K20	K21-K25	Off-list
1	46.8	41.6	1.7	6.5	2.5	1.1
2	61.3	18.1	9.3	1.9	3.7	5.4
3	36.3	29	2.9	5.7	4.2	22.1
4	87.1	7.1	2	1.1	0.2	2.2
5	88.4	4.2	1.5	0.3	0	5.4
10 th grade						
Prompt	K1-K5	K6-K10	K11-K15	K16-K20	K21-K25	Off-list
1	58.5	31.6	2.1	4.9	1.4	1.4
2	52.6	20.5	12.7	2	3.4	8.8
3	35.4	31.7	4.5	2.8	5	20.6
4	87.2	5.3	3.5	0.6	0.4	2.8
5	83.2	8.9	0.5	0	0.5	7
11 th grade						
Prompt	K1-K5	K6-K10	K11-K15	K16-K20	K21-K25	Off-list
1	51	37.9	1.4	5.1	2.6	2.2
2	55.6	18.2	8.6	1.9	4	11.5
3	38.8	29.2	4	4.1	4	19.6
4	81.3	6.4	1.8	0.9	0.2	9.2
5	88.9	4.2	0.3	0.3	0.9	5.5

Note. Prompt 1 = *Animals*, Prompt 2 = *Food & Drink*, Prompt 3 = *Sport and Physical Activities*, Prompt 4 = *Environment & Climate*, Prompt 5 = *Economy & Money*.

Regarding French, similar results were obtained. The categories *Environment & Climate* and *Economy & Money* produced a higher percentage of words in the K-1 to K-5 lists than the other categories, as was the case in the English lists. As shown in the table above, this band accounted for 87.1% of tokens in 9th grade, 87.2% of tokens in 10th grade and 81.3% of tokens in 11th grade for *Environment & Climate*, and for 88.4% of tokens in 9th grade, 83.2% of tokens in 10th grade and 88.9% of tokens in 11th grade. This again indicates that, as well as producing fewer tokens in these domains, the types that were produced were among the most frequent in the French v.5 Corpora. Again, there was also a high percentage of off-list words in the category *Sport & Physical Activities*, making up 22.11%, 20.65% and 19.67% of the tokens in 9th, 10th and 11th grade, respectively. Unlike in the case of English, a similar observation was made for the prompt *Food & Drink* particularly at higher levels, where off-list words accounted for 5.4%, 8.8% and 11.5% of the tokens in 9th, 10th and 11th grade, respectively. In some cases, this can be attributed to the presence of single-word units in English which are multiword units in French, such as “pomme de terre” [potatoes]. In other cases, a term such as “épinard” [spinach] was

classified as off-list in the French corpora while in the English corpora it appeared as a relatively frequent word at the K7 level.

Regarding the cross-sectional and longitudinal qualitative analysis, very few differences were observed across the three grades. The most frequent first word for each prompt did not vary much across grades in the majority of prompts in either language, with slight differences found only for the prompt *Sport & Physical Activities* in French in 10th grade and the prompt *Economy & Money* across the three grades in both languages. In the former, the most frequent first word for 10th grade participants was “basketball” in French, as compared to “football” at all grades in English and 9th and 11th in French. In the latter, the most frequent first word in 9th, 10th and 11th grade respectively was “dollar”, “economy” and “money” in English and “euro”, “economy” and “money” in French. However, despite these slight differences, there is no indication that CLIL subject has influenced the vocabulary produced: in all cases students produce either a currency or a word that is contained in the prompt title. The ranking of the most and least productive prompts revealed practically no differences across the grades for English, while some differences were again found in 10th grade in French. The content-related prompt *Environment & Climate* was the most productive prompt in 9th and 11th grade in French, whereas it ranked third for 10th grade students. The cognate analysis also indicated little variation in English but greater variety in French. In English, percentages did not vary more than 2%, with the exception of the prompt *Economy & Money: Animals* (26%, 25%, 27%), *Food & Drink* (46%, 44%, 45%), *Sport & Physical Activity* (39%, 38%, 39%), *Environment & Climate* (20%, 21%, 21%), *Economy & Money* (42%, 49%, 45%). In French, greater differences were observed across the prompts, with the exception of *Food & Drink* which remained relatively constant across the three grades: *Animals* (26%, 12%, 15%), *Food & Drink* (39%, 39%, 41%), *Sport & Physical Activity* (83%, 76%, 78%), *Environment & Climate* (34%, 31%, 27%), *Economy & Money* (75%, 69%, 69%). It should be noted that in French, 9th grade students’ production in French contained the highest percentage of English cognates, indicating that they may rely more on English than the older students. As noted above, the content-relevant prompt *Environment & Climate* was the most productive prompt in the French LAT for 9th and 11th grade students, but not for 10th grade students. It is interesting to note that in 10th grade, the prompt with the highest percentage of cognates *Sport & Physical Activities*, was in fact that which ranked as the most productive prompt for French. Thus, productivity may well have been influenced by the presence of English cognates, leading to the difference between 10th

grade and 9th and 11th grade. Lexical sophistication in terms of the ranking of non-shared words and the number of infrequent words in the production of each prompt also indicated no real difference across the three grades in either language. The prompt *Economy & Money* contained the highest number of infrequent words and *Animals* contained the fewest number of non-shared words at all three grades, and a higher number of non-shared words across was observed in content-relevant prompts at all three grades in both languages. No clear differences were found the number of infrequent words across grades in either language.

In summary, the analysis of most frequent first words for each prompt indicated a large degree of similarity in each language. Some exceptions included the prompt *Sport & Physical Activities* in 10th grade, where the most common first word in English was “football”, but “basketball” in French, and in *Economy & Money*, where the most common first words for 9th grade students was “dollar” in English and “euro” in French. The analysis of the ranking of the most and least productive prompts revealed a great degree of variation in the productivity of each prompt in English as compared to French. Of particular interest was that English content-relevant prompts were the least productive at all three grades in the English LAT, whereas the French content-relevant prompt was found to be the most productive for 9th grade and 11th grade students in the French LAT. Thus, although they studied physical education and economics through English, students did not produce more words for these prompts than other more productive ones such as *Food & Drink* or *Animals* in the English LAT, whereas studying geography and history through French may have provided 9th and 11th grade students with enough lexical items to exceed the number produced for more productive prompts. Regarding the analysis of non-shared words, there were clear similarities in both languages in the ranking, and content-relevant prompts generally revealed a higher number of non-shared words for both English and French. However, students also generally produced a higher percentage of non-shared words in English than in French, with the exception of *Animals* and *Food & Drink* in 9th and 11th grade and *Environment & Climate* in 11th grade. Finally, regarding the number of infrequent words produced in each language, despite the use of different corpora for each language, results were quite similar in that the less common semantic categories, *Environment & Climate* and *Economy & Money*, contained a higher percentage of words in the K-1 to K-5 lists than the other categories. In addition, a high percentage of off-list words were found in the prompt *Sport & Physical Activities* in both languages and in the prompt *Food & Drink* in French, likely due to the presence of

multiword units. The cross-sectional and longitudinal analysis revealed very few differences across the three grades, though some minor differences were found for the most frequent first word in the prompt *Economy & Money* in each language, and the ranking of the most and least productive prompts and percentage of English cognates in French.

8.2. Target Language Influence on Language Learning Motivation

The second research question aimed to determine, firstly, whether there was a difference between the participants' language learning motivation towards English as compared to French and, secondly, if there was a relationship between the participants' LA, language level and their language learning motivation in each language.

8.2.1. Quantitative Differences in Motivation towards English and French

Research question 2.1 asked whether there were quantitative differences between the participants' language learning motivation towards English as compared to French at each testing period. The descriptive statistics for the language learning motivation in each language, as shown in Table 8.13, indicated that participants reported higher language learning motivation in English both overall and in all individual categories at all grades. In order to determine if these differences in language learning motivation were statistically significant, Wilcoxon signed-rank tests were conducted to compare the participants' language learning motivation towards English as compared to French at each testing period (Table 8.14). These results are presented visually in Figure 8.4.

Table 8.13*Descriptive Statistics for Language Learning Motivation in English and French*

9 th grade								
English (<i>n</i> = 39)				French (<i>n</i> = 41)				
Min	Max	<i>Mdn</i>	<i>SD</i>	Category	Min	Max	<i>Mdn</i>	<i>SD</i>
1.80	5.00	4.00	.85	Ideal L2 Self	1.20	5.00	2.80	.98
1.57	3.71	2.57	.58	The “Ought to” Self	1.14	3.57	2.57	.63
1.80	5.00	4.00	.92	Language Anxiety	1.00	5.00	3.80	1.11
2.33	5.00	3.83	.73	Interest in FLs	1.00	5.00	3.17	.95
2.25	5.00	3.75	.69	L2 Self Confidence	1.00	4.50	3.25	.72
1.00	5.00	4.00	.87	Instrumentality: Prevention	1.20	4.80	2.80	.86
1.50	5.00	4.17	.74	Instrumentality: Promotion	1.00	4.83	3.17	.96
1.50	4.71	3.36	.82	Attitude towards Learning	1.25	4.50	2.38	.81
2.20	5.00	3.80	.62	Intended Learning Effort	1.00	5.00	3.00	.90
2.86	4.51	3.65	.47	Mean Motivation	1.77	4.29	2.92	.53
10 th grade								
English (<i>n</i> = 41)				French (<i>n</i> = 41)				
Min	Max	<i>Mdn</i>	<i>SD</i>	Category	Min	Max	<i>Mdn</i>	<i>SD</i>
2.60	5.00	4.00	.63	Ideal L2 Self	1.00	4.40	3.00	.88
2.14	3.86	3.00	.46	The “Ought to” Self	1.57	3.71	2.71	.50
1.60	5.00	3.80	.91	Language Anxiety	1.20	5.00	3.60	.95
2.17	4.83	3.83	.60	Interest in FLs	1.67	4.50	3.33	.64
2.00	5.00	3.75	.81	L2 Self Confidence	1.00	5.00	3.25	.84
2.40	5.00	4.00	.67	Instrumentality: Prevention	1.20	4.20	3.00	.68
3.00	5.00	4.17	.48	Instrumentality: Promotion	1.17	4.83	3.50	.79
2.11	4.40	3.60	.53	Attitude towards Learning	1.00	4.00	3.00	.81
2.40	4.80	3.80	.54	Intended Learning Effort	1.20	4.40	3.00	.79
2.93	4.42	3.77	.41	Mean Motivation	1.61	4.04	3.19	.56
11 th grade								
English (<i>n</i> = 40)				French (<i>n</i> = 36)				
Min	Max	<i>Mdn</i>	<i>SD</i>	Category	Min	Max	<i>Mdn</i>	<i>SD</i>
2.60	5.00	4.20	.69	Ideal L2 Self	1.20	4.40	3.00	.90
1.71	4.43	3.00	.60	The “Ought to” Self	1.14	3.86	2.57	.60
2.00	5.00	4.10	.86	Language Anxiety	1.80	5.00	3.80	.96
2.50	5.00	3.83	.62	Interest in FLs	1.83	4.50	3.50	.69
1.75	5.00	3.75	.81	L2 Self Confidence	1.25	5.00	3.25	.74
2.60	5.00	4.00	.64	Instrumentality: Prevention	1.00	4.40	2.70	.92
2.67	5.00	4.08	.57	Instrumentality: Promotion	1.67	5.00	3.25	.89
1.75	4.60	3.37	.66	Attitude towards Learning	1.38	4.00	3.06	.63
1.80	4.80	3.80	.64	Intended Learning Effort	1.00	4.00	3.00	.79
2.77	4.59	3.79	.41	Mean Motivation	1.85	3.93	3.11	.52

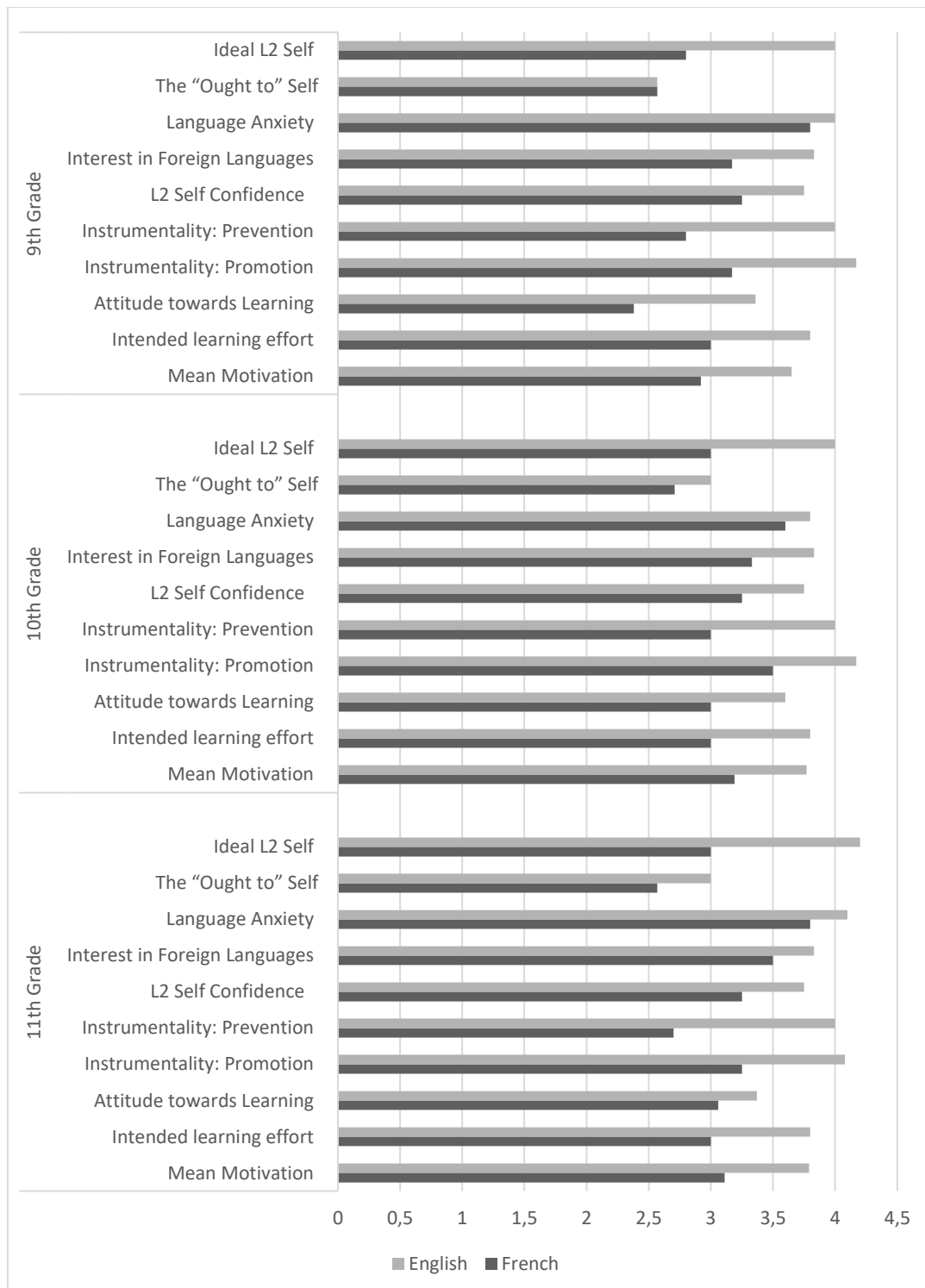
Table 8.14*Results of Wilcoxon Signed-rank Tests Comparing MFQ in English and French*

9 th Grade						
Category	Mdn		SD		z	p
	EN	FR	EN	FR		
Ideal L2 Self	4.00	2.80	.85	.98	-4.91	<.001
The “Ought to” Self	2.57	2.57	.58	.63	-1.83	.066
Language Anxiety	4.00	3.80	.92	1.11	-2.60	.009
Interest in Foreign Languages	3.83	3.17	.73	.95	-4.26	<.001
L2 Self Confidence	3.75	3.25	.69	.72	-3.02	.002
Instrumentality: Prevention	4.00	2.80	.87	.86	-5.11	<.001
Instrumentality: Promotion	4.17	3.17	.74	.96	-4.91	<.001
Attitude towards Learning	3.36	2.38	.82	.81	-4.19	<.001
Intended Learning Effort	3.80	3.00	.62	.90	-5.03	<.001
Mean Motivation	3.65	2.92	.47	.53	-5.35	<.001
10 th Grade						
Category	Mdn		SD		z	p
	EN	FR	EN	FR		
Ideal L2 Self	4.00	3.00	.63	.88	-5.18	<.001
The “Ought to” Self	3.00	2.71	.46	.50	-4.65	<.001
Language Anxiety	3.80	3.60	.91	.95	-1.73	.083
Interest in Foreign Languages	3.83	3.33	.60	.64	-3.11	.002
L2 Self Confidence	3.75	3.25	.81	.84	-1.75	.080
Instrumentality: Prevention	4.00	3.00	.67	.68	-4.68	<.001
Instrumentality: Promotion	4.17	3.50	.48	.79	-5.12	<.001
Attitude towards Learning	3.60	3.00	.53	.81	-4.76	<.001
Intended Learning Effort	3.80	3.00	.54	.79	-4.88	<.001
Mean Motivation	3.77	3.19	.41	.56	-4.95	<.001
11 th Grade						
Category	Mdn		SD		z	p
	EN	FR	EN	FR		
Ideal L2 Self	4.20	3.00	.69	.90	-4.59	<.001
The “Ought to” Self	3.00	2.57	.60	.60	-3.36	.001
Language Anxiety	4.10	3.80	.86	.96	-2.32	.020
Interest in Foreign Languages	3.83	3.50	.62	.69	-2.54	.011
L2 Self Confidence	3.75	3.25	.81	.74	-1.58	.113
Instrumentality: Prevention	4.00	2.70	.64	.92	-4.94	<.001
Instrumentality: Promotion	4.08	3.25	.57	.89	-4.83	<.001
Attitude towards Learning	3.37	3.06	.66	.63	-1.97	.049
Intended Learning Effort	3.80	3.00	.64	.79	-4.65	<.001
Mean Motivation	3.79	3.11	.41	.52	-4.80	<.001

Note. EN = English, FR = French. 9th grade: English ($n = 31$), French ($n = 40$); 10th grade English ($n = 41$), French ($n = 41$), 11th grade English ($n = 40$), French ($n = 36$).

Figure 8.4

Motivation in English and French



As shown, at all three grades, results revealed statistically significant differences in the participants' overall language learning motivation, and nearly all categories, with participants' indicating a higher level of motivation towards English than to French in all

cases. Statistically significant differences between the motivation for each language were found in all categories except *The “Ought to” Self* in 9th grade, in all categories except *Language Anxiety* and *L2 Self Confidence* in 10th grade, and in all categories except *L2 Self Confidence* in 11th grade. To this effect, there were statistically significant differences in six out of the nine categories across all three grades: *Ideal L2 Self*, *Interest in Foreign Languages*, *Instrumentality: Prevention*, *Instrumentality: Promotion*, *Attitude towards Learning* and *Intended Learning Effort*. These results indicate that in English as compared to French, participants can better see themselves as the L2 user they wish to be; are more interested in the language itself; are more instrumentality motivated, seeing not learning English as preventing their future success and having English as promoting their future success; have better attitudes towards learning English than French; and finally, have a higher intended learning effort in English. On the other hand, though means were higher for English in all cases, no statistically significant differences were found for *The “Ought to” Self* in 9th grade, *Language Anxiety* in 10th grade, and *L2 Self Confidence* in 10th and 11th grade. This suggests that for the younger participants there is no key difference in the way they are externally motivated to learn English as compared to French, while the older participants’ self-confidence or anxiety may be similar regardless of the language at hand. In other words, while 10th and 11th grade students report a noticeable difference in how they perceive English and French in terms of categories such as instrumentality or interest in the language, this perception may not affect their own self-confidence and anxiety, and so if they lack confidence in one language may also lack it in the other.

In order to determine whether there were cross-sectional (9th and 10th grade) and longitudinal differences (10th and 11th grade) between the participants’ language learning motivation towards English and French, Mann-Whitney U tests and Wilcoxon signed-rank tests, respectively, were carried out. Firstly, regarding cross-sectional differences in English (Table 8.15) and French (Table 8.16), results indicated that there were very few differences between students in 9th and 10th grade, with no statistically significant differences in overall motivation in either language and only in the category *The “Ought to” Self* in English. This indicates that students in 9th and 10th grade are equally motivated to learn English and French overall and in the majority of the motivation categories, though students in 10th grade place greater importance on external sources of motivation towards English. These results are presented visually in Figure 8.5.

Table 8.15*Cross-Sectional Differences in English Language Learning Motivation*

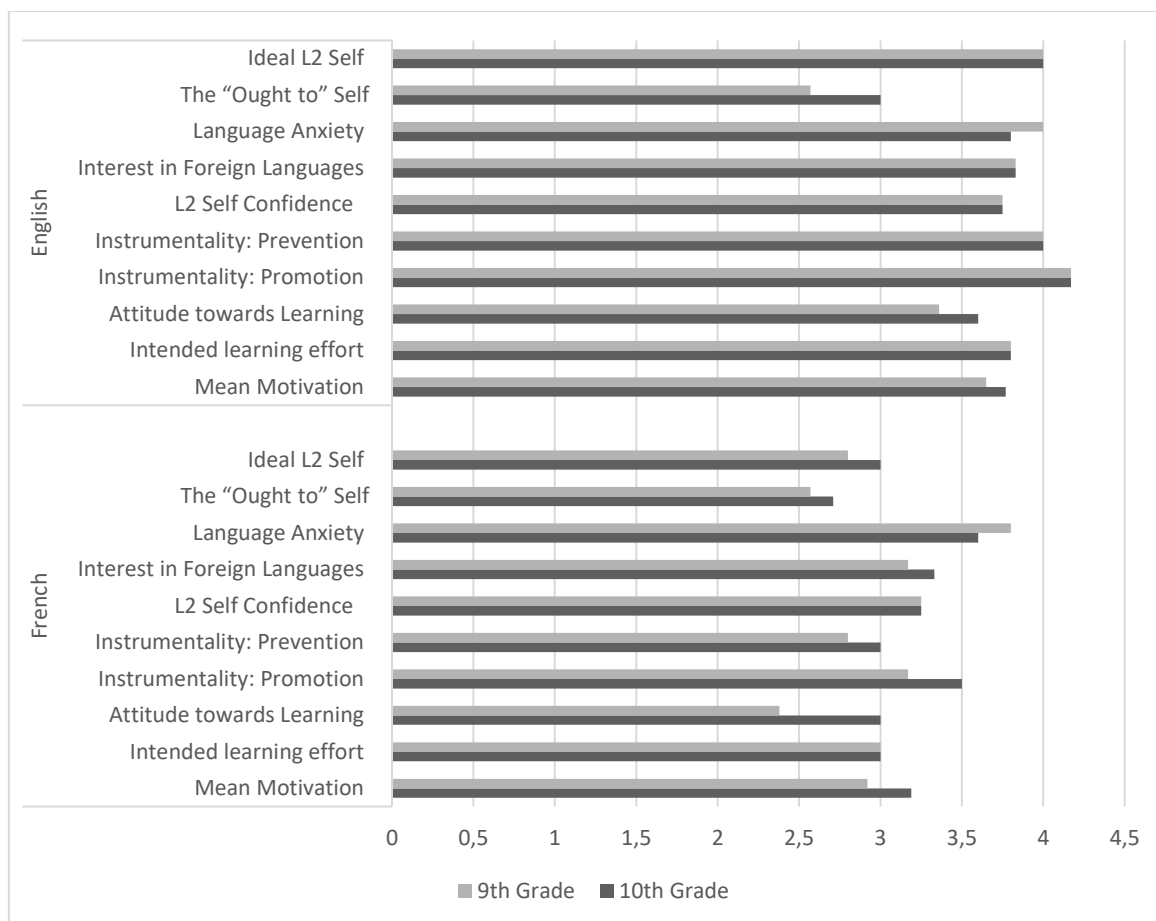
Category	Mdn		SD		z	p
	9 th	10 th	9 th	10 th		
Ideal L2 Self	4.00	4.00	.85	.63	-.10	.919
The "Ought to" Self	2.57	3.00	.58	.46	-2.85	.004
Language Anxiety	4.00	3.80	.92	.91	-.50	.612
Interest in Foreign Languages	3.83	3.83	.73	.60	-.80	.422
L2 Self Confidence	3.75	3.75	.69	.81	-1.27	.204
Instrumentality: Prevention	4.00	4.00	.87	.67	-.59	.555
Instrumentality: Promotion	4.17	4.17	.74	.48	-.45	.653
Attitude towards Learning	3.36	3.60	.82	.53	-.77	.441
Intended Learning Effort	3.80	3.80	.62	.54	-1.55	.120
Mean English Motivation	3.65	3.77	.47	.41	-.22	.821

Note. 9th grade, $n = 39$; 10th grade, $n = 41$

Table 8.16*Cross-Sectional Differences in French Language Learning Motivation*

Category	Mdn		SD		z	p
	9 th	10 th	9 th	10 th		
Ideal L2 Self	2.80	3.00	.98	.88	-1.20	.227
The "Ought to" Self	2.57	2.71	.63	.50	-1.62	.105
Language Anxiety	3.80	3.60	1.11	.95	-.20	.842
Interest in Foreign Languages	3.17	3.33	.95	.64	-1.62	.103
L2 Self Confidence	3.25	3.25	.72	.84	-.30	.758
Instrumentality: Prevention	2.80	3.00	.86	.68	-1.57	.116
Instrumentality: Promotion	3.17	3.50	.96	.79	-1.54	.123
Attitude towards Learning	2.38	3.00	.81	.81	-.97	.332
Intended Learning Effort	3.00	3.00	.90	.79	-.54	.583
Mean French Motivation	2.92	3.18	.53	.56	-1.76	.077

Note. 9th grade, $n = 41$; 10th grade, $n = 41$

Figure 8.5*Cross-Sectional Differences in Motivation*

Secondly, regarding longitudinal differences in English (Table 8.17) and French (Table 8.18), similar results were obtained, with students' motivation towards English and French remaining relatively constant from 10th to 11th grade. No statistically significant differences were found for overall motivation in either language or in the majority of the individual categories. However, with regards to English, statistically significant differences were found in the category *Attitude towards Learning*, with 11th grade students reporting lower motivation. This indicates that, perhaps following the changes to their exposure to English given that the students generally no longer took English CLIL classes, their attitude was not as positive towards learning English as it had been in 10th grade. With regards to French, statistically significant differences were found in two categories: *Instrumentality: Prevention* and *Instrumentality: Promotion*. In both cases, 11th grade students reported lower motivation in these two categories. This suggests that these students decreased the extent to which they saw not having French as a hindrance to their future success and the extent to which they saw having it as a way to

promote their future success. This again could be attributed to the elimination of CLIL teaching in French in 11th grade, as given that the students received less exposure and French became less of a priority, they could consequently view French as less important to their futures. These results are presented visually in Figure 8.6.

Table 8.17*Longitudinal Differences in Longitudinal English Language Learning Motivation*

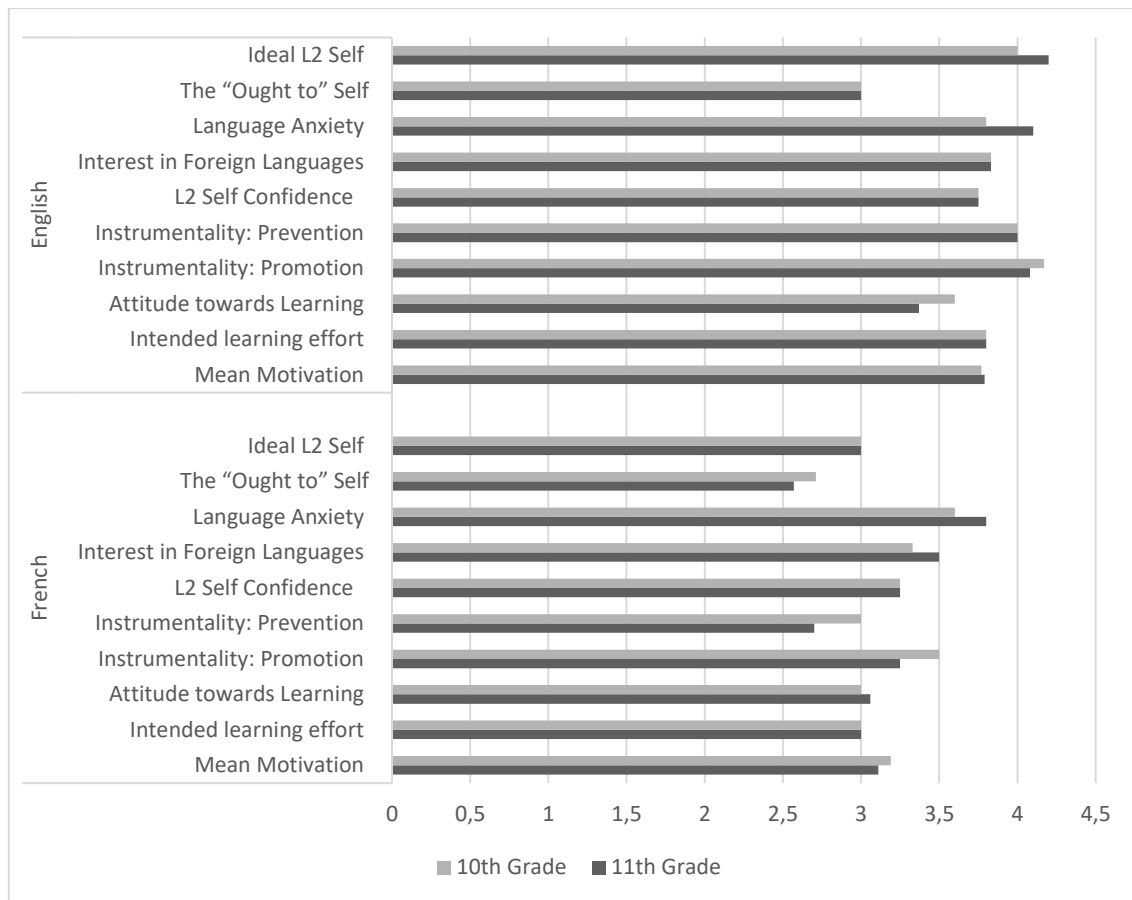
Category	Mdn		SD		z	p
	10 th	11 th	10 th	11 th		
Ideal L2 Self	4.00	4.20	.63	.69	-.59	.549
The “Ought to” Self	3.00	3.00	.46	.60	-1.52	.127
Language Anxiety	3.80	4.10	.91	.86	-.15	.878
Interest in Foreign Languages	3.83	3.83	.60	.62	-1.16	.246
L2 Self Confidence	3.75	3.75	.81	.81	-.29	.766
Instrumentality: Prevention	4.00	4.00	.67	.64	-.55	.582
Instrumentality: Promotion	4.17	4.08	.48	.57	-.36	.714
Attitude towards Learning	3.60	3.37	.53	.66	-3.27	.001
Intended Learning Effort	3.80	3.80	.54	.64	-1.20	.230
Mean English Motivation	3.77	3.79	.41	.41	-1.68	.092

Note. 10th grade, $n = 41$; 11th grade, $n = 40$

Table 8.18*Longitudinal Differences in French Language Learning Motivation*

Category	Mdn		SD		z	p
	10 th	11 th	10 th	11 th		
Ideal L2 Self	3.00	3.00	.88	.90	-.84	.400
The “Ought to” Self	2.71	2.57	.50	.60	-.93	.350
Language Anxiety	3.60	3.80	.95	.96	-.40	.684
Interest in Foreign Languages	3.33	3.50	.64	.69	-.66	.508
L2 Self Confidence	3.25	3.25	.84	.74	-.49	.617
Instrumentality: Prevention	3.00	2.70	.68	.92	-2.55	.011
Instrumentality: Promotion	3.50	3.25	.79	.89	-2.27	.023
Attitude towards Learning	3.00	3.06	.81	.63	-.11	.905
Intended Learning Effort	3.00	3.00	.79	.79	-1.89	.058
Mean French Motivation	3.19	3.11	.56	.52	-1.12	.262

Note. 10th grade, $n = 41$; 11th grade, $n = 36$

Figure 8.6*Longitudinal Differences in Motivation*

In summary, clear differences were found between students' motivation towards English and French across all three grades. Statistically significant differences were found in the participants' overall language learning motivation for each language, and nearly all categories, with participants' indicating a higher level of motivation towards English than to French in all cases. Some exceptions were *The "Ought to" Self* in 9th grade, *Language Anxiety* and *L2 Self Confidence* in 10th grade, and *L2 Self Confidence* in 11th grade. The cross-sectional analysis revealed few differences between students in 9th and 10th grade, with statistically significant differences found only in *The "Ought to" Self* in English. The longitudinal analysis similarly also found few differences in students from 10th to 11th grade. Statistically significant differences, however, were found in the category *Attitude towards Learning* in English, with 11th grade students reporting lower motivation, and in the categories *Instrumentality: Prevention* and *Instrumentality: Promotion* in French, with 11th grade students again reporting lower motivation.

8.2.2. *Lexical Availability, Language Level and Language Learning Motivation*

In order to address research question 2.2, Spearman's rank order correlations, given that the data for motivation was ordinal, were used to determine whether there was a relationship between the results of the participants' LAT and their responses in the MFQ, and their language level and their responses in the MFQ. This was done first for English and then for French. Regarding English, as shown in Table 8.19, results showed a statistically significant moderate positive correlation between the *Mean_LAT* and *Mean_MFQ* in 9th grade ($r(37) = .37, p = .018$), 10th grade ($r(39) = .43, p = .005$) and 11th grade ($r(38) = .42, p = .007$). This indicates that for students in all grades, there was a relationship between the number of words produced by students in English and their English language learning motivation, with more motivated students retrieving a higher number of words. With regards to the individual categories, statistically significant correlations were found in six out of the nine categories at one or more levels. Firstly, statistically significant moderate positive correlations were found between the *Mean_LAT* and *Language Anxiety* ($r(39) = .31, p = .042$) only in 10th grade. This suggests that 10th grade students with lower language anxiety produced more words in the English LAT. Secondly, in both 10th and 11th grade, who were notably largely the same cohort of students, statistically significant moderate positive correlations were found between the *Mean_LAT* and *Ideal L2 Self* (10th grade: $r(39) = .39, p = .010$; 11th grade: $r(38) = .37, p = .018$), and between the *Mean_LAT* and *Interest in Foreign Languages* (10th grade: $r(39) = .39, p = .010$; 11th grade: $r(38) = .33, p = .037$). This indicates that students in these groups who retrieved a higher number of words in the English LAT were better able to visualise themselves as the L2 user they wished to be and reported higher interest in learning English. In 9th grade and 11th grade, there was a statistically significant moderate correlation between the *Mean_LAT* and *Intended Learning Effort* (9th grade: $r(37) = .36, p = .023$); 11th grade: $r(38) = .39, p = .012$), indicating that 9th and 11th grade students who retrieved a higher number of words in the English LAT also reported making more of an effort in learning English. Finally, at all three levels there were statistically significant moderate positive correlations between the *Mean_LAT* and *L2 Self Confidence* (9th grade: $r(37) = .32, p = .047$; 10th grade: $r(39) = .38, p = .014$; 11th grade: $r(38) = .45, p = .003$), and between *Mean_LAT* and *Instrumentality: Prevention* (9th grade: $r(37) = .40, p = .010$; 10th grade: $r(39) = .39, p = .010$; 11th grade: $r(38) = .31, p = .045$). These results show that the participants who retrieved the highest number of words were

those who were more confident with the L2 and who saw not learning English as preventing their future success.

Table 8.19

Spearman Rank Correlations between Mean LAT and MFQ Categories in English

	9 th Grade Mean LAT	10 th Grade Mean LAT	11 th Grade Mean LAT
Ideal L2 Self	.21	.39*	.37*
The “Ought to” Self	-.23	-.29	-.19
Language Anxiety	.19	.31*	.25
Interest in Foreign Languages	.25	.39*	.33*
L2 Self Confidence	.32*	.38*	.45**
Instrumentality: Prevention	.40*	.39*	.31*
Instrumentality: Promotion	.20	.04	.11
Attitude towards Learning	.02	.28	.13
Intended Learning Effort	.36*	.28	.39*
Mean MFQ English	.37*	.43**	.42**

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

No statistically significant correlations were found at any level between the *Mean_LAT* and the remaining three categories: *The “Ought to” Self*, *Instrumentality: Promotion* or *Attitude towards Learning*. This implies that there was no relationship between the number of words retrieved by students in English and the extent to which they were motivated by external sources, saw English as promoting their future success or how positive or negative their attitude was towards the language.

With regards to language level, results also showed a significant moderate positive correlation between the results of the C-Test and the MFQ in 10th grade ($r(39) = .37, p = .015$) and 11th grade ($r(38) = .35, p = .025$). However, the same was not found in 9th grade: $r(37) = .24, p = .126$. This indicates that, in the case of English, older students who performed better on the C-test also reported higher motivation towards English overall, though there was no statistically significant relationship in the case of the younger students.

Regarding French, as shown in Table 8.20, results showed that there was a statistically significant moderate correlation between the *Mean_LAT* and *Mean_MFQ*, in 10th grade ($r(39) = .33, p = .031$), while this was not the case in 9th grade ($r(39) = .24, p = .124$) and 11th grade ($r(34) = .11, p = .491$). This indicates that students in 10th grade who produced a higher number of words on the French LAT reported higher motivation

towards French overall, whereas there was no such relationship in 9th or 11th grade.

With regards to the individual categories, statistically significant correlations were found in four out of the nine categories at one or more levels. Firstly, in 10th grade only, a statistically significant high positive correlation was found between the *Mean_LAT* and *Ideal L2 Self* ($r(39) = .52, p = < .001$) and between the *Mean_LAT* and *Instrumentality: Promotion* ($r(39) = .52, p = < .001$). This suggests that 10th grade students who retrieved a higher number of words in the French LAT were better able to visualise themselves as the French language user they wished to be and saw French as important in promoting their future success. Secondly, in both 9th and 10th grade, when students took content classes through French, a statistically significant moderate negative correlation was found between the *Mean_LAT* and *The “Ought to” Self* (9th grade: $r(39) = -.35, p = .024$; 10th grade: $r(39) = -.42, p = .006$) and a statistically significant moderate positive correlation was found between the *Mean_LAT* and *L2 Self Confidence* (9th grade: $r(39) = .50, p = .001$; 10th grade: $r(39) = .43, p = .005$). This indicates that students in these groups who retrieved a higher number of words in the French LAT were less motivated by external sources and had higher self-confidence in French.

Table 8.20

Spearman Rank Correlations between LAT and MFQ Categories in French

	9 th Grade Mean LAT	10 th Grade Mean LAT	11 th Grade Mean LAT
Ideal L2 Self	.17	.52**	.02
The “Ought to” Self	-.35*	-.42*	-.12
Language Anxiety	.25	.25	.11
Interest in Foreign Languages	.17	.21	.02
L2 Self Confidence	.50*	.43**	.28
Instrumentality: Prevention	.28	.23	.18
Instrumentality: Promotion	.26	.52*	.01
Attitude towards Learning	-.10	.27	.13
Intended Learning Effort	.00	.29	-.06
Mean MFQ French	.24	.33*	.11

*. Correlation is significant at the 0.05 level (2-tailed).

** . Correlation is significant at the 0.01 level (2-tailed).

No categories showed statistically significant positive or negative correlations across all three levels and no statistically significant correlations were found at any level between the *Mean_LAT* and the remaining five categories: *Language Anxiety*, *Interest in Foreign Languages*, *Instrumentality: Prevention*, *Attitude towards Learning* and *Intended*

learning effort. This implies that there was no relationship between the number of words retrieved by students in French and how anxious they were towards the language, how interested they were in the language, the extent to which they believed not having French would prevent their future success, how positive or negative their attitude was towards the language, or the amount of effort they made to learn French. In addition, it should be noted that no statistically significant correlations were found at any level for 11th grade students, suggesting that there was no relationship between these students' performance on the French LAT and any area of their motivation towards learning French.

Regarding language level, results also showed that there was no statistically significant correlation between the results of the C-Test and the MFQ in French in 9th grade ($r(39) = .26, p = .097$), 10th grade ($r(39) = .22, p = .166$), or 11th grade ($r(34) = -.06, p = .718$). This indicates that, unlike in the case of English, students who performed better on the French C-test did not necessarily have higher motivation towards French, nor did those who received a lower score report lower motivation.

In summary, clear differences were found between the relationship between LA, motivation and language level in each language. In English, significant moderate positive correlations were found between the *Mean_LAT* and *Mean_MFQ* at all levels and between the C-test and the *Mean_MFQ* and in 10th and 11th grade, but not in 9th grade. In French, a significant moderate positive correlation was found between the *Mean_LAT* and *Mean_MFQ* in 10th grade, but not in 9th or 11th grade, and no statistically significant correlation was found between the results of the C-Test and the *Mean_MFQ* at any level. In addition, with regards to the individual categories, a number of differences arose across languages and categories. As shown in Table 8.21, no relationship was found between the *Mean_LAT* in English and the “*Ought to*” *Self, Instrumentality: Promotion* or *Attitude towards Learning* at any level. Meanwhile, in French, no relationship was found between the *Mean_LAT* and *Language Anxiety, Interest in Foreign Languages, Instrumentality: Prevention, Attitude towards Learning, Intended learning effort*. This suggests that, in English, there is no relationship between LA and external sources of motivation, how the students view English as promoting their future success or their attitude towards the language. In French, on the other hand, there appears to be no relationship between LA and how anxious the students are, how interested they are in French, how they see not having French as preventing their future success, their attitude towards the language, or the amount of effort they made to learn French.

Table 8.21*Summary of Non-significant Correlations in English and French*

English		French	
Mean_LAT	The “Ought to” Self	Mean_LAT	Language Anxiety
	Instrumentality: Promotion		Interest in FLs
	Attitude towards Learning		Instrumentality: Prevention
	Attitude towards Learning		
	Intended Learning Effort		

8.3. Gender

The third research question aimed to determine whether there were quantitative (RQ3.1) and qualitative (RQ3.2) differences between the male groups and the female groups in each grade at each testing period with regard to LA and whether there were quantitative differences between the male groups and the female groups with regard to motivation (RQ3.3).

8.3.1. *Quantitative Differences in Gender and LA*

In order to address research question 3.1, two groups of analyses were carried out. The first investigated the words retrieved by male and female students in English as compared to French at each testing period, that is, within gender and across languages (RQ3.1.1), while the second addressed the words retrieved by male and female participants in English and in French at each testing period, that is, across gender and within languages (RQ3.1.2).

8.3.1.1. Quantitative Differences in Gender and LA across Languages. Research question 3.1.1 asked whether there were quantitative differences in the words retrieved by male and female participants in English as compared to French in each grade at each testing period. The descriptive statistics for the overall LAT for each language, as shown in Table 8.22, indicated that both male and female participants produced a higher number of words in English than they did in French in all three grades, both in the overall LAT and in each of the five individual prompts. The normality of each prompt and the overall LAT in each language was then assessed for each grade and gender. The results are summarised in Table 8.23. As shown, data were normally distributed in all cases with the exception of the prompt *Animals* in English for 9th grade male students, *Sport & Physical Activities* in French for 9th grade male students, *Animals* in French for 9th grade female students, and *Environment & Climate* in English for 11th grade male students.

Table 8.22*Descriptive Statistics for Lexical Availability within Gender across Languages*

9 th grade: Male								
English (<i>n</i> = 16)					French (<i>n</i> = 19)			
Min	Max	<i>M</i>	<i>SD</i>	Prompt	Min	Max	<i>M</i>	<i>SD</i>
1	31	16.94	8.31	Animals	2	18	7.47	3.53
0	35	13.88	8.17	Food & Drink	1	15	7.05	3.58
3	29	14.06	7.10	Sport & PA	0	15	6.95	3.70
4	33	14.19	8.47	Env & Climate	1	19	8.63	5.37
0	24	7.69	6.61	Economy & Money	0	14	4.95	3.89
21	151	66.75	34.52	Mean LAT	10	70	35.05	14.63
9 th grade: Female								
English (<i>n</i> = 23)					French (<i>n</i> = 23)			
Min	Max	<i>M</i>	<i>SD</i>	Prompt	Min	Max	<i>M</i>	<i>SD</i>
13	32	20.57	5.86	Animals	4	29	12.70	5.98
8	32	23.87	6.31	Food & Drink	5	25	12.23	4.73
4	24	15.00	5.04	Sport & PA	6	18	11.68	2.93
9	34	19.48	6.29	Env & Climate	2	23	11.83	4.68
1	20	12.35	4.83	Economy & Money	3	18	10.26	4.57
38	129	91.26	24.80	Mean LAT	32	109	57.65	17.46
10 th grade: Male								
English (<i>n</i> = 15)				English (<i>n</i> = 15)	English (<i>n</i> = 15)			
Max	Max	Max	Max	Prompt	Max	Max	Max	Max
12	28	17.73	4.49	Animals	2	12	8.40	2.77
12	29	19.87	4.73	Food & Drink	2	16	8.07	4.31
10	20	14.93	2.76	Sport & PA	9	16	11.47	2.03
2	26	15.27	6.19	Env & Climate	3	16	7.60	3.66
4	18	11.60	4.71	Economy & Money	3	16	10.67	3.71
46	107	79.40	17.06	Mean LAT	25	70	46.20	12.90

Table 8.22 (continued)

10 th grade: Female								
English (<i>n</i> = 26)					French (<i>n</i> = 26)			
Min	Max	<i>M</i>	<i>SD</i>	Prompt	Min	Max	<i>M</i>	<i>SD</i>
10	28	18.50	5.08	Animals	2	19	8.62	3.87
15	35	22.62	5.42	Food & Drink	3	23	12.38	5.21
9	27	17.58	5.13	Sport & Physical Activities	6	17	11.08	2.85
12	31	20.54	4.79	Environment & Climate	5	23	12.04	4.60
10	26	15.31	4.36	Economy & Money	2	17	10.27	3.82
62	146	94.53	19.81	Mean LAT	27	83	54.38	15.97
11 th grade: Male								
English (<i>n</i> = 19)					French (<i>n</i> = 15)			
Min	Max	<i>M</i>	<i>SD</i>	Prompt	Min	Max	<i>M</i>	<i>SD</i>
7	29	18.05	5.47	Animals	4	16	9.27	3.93
8	28	18.11	6.42	Food & Drink	1	16	7.73	4.20
9	21	15.32	4.37	Sport & PA	5	14	9.87	3.33
4	26	14.47	6.52	Env & Climate	2	16	7.93	4.57
2	20	12.26	4.74	Economy & Money	0	12	7.33	3.37
48	112	78.21	20.81	Mean LAT	16	68	42.13	15.70
11 th grade: Female								
English (<i>n</i> = 21)					French (<i>n</i> = 21)			
Min	Max	<i>M</i>	<i>SD</i>	Prompt	Min	Max	<i>M</i>	<i>SD</i>
12	30	21.71	5.36	Animals	6	18	12.33	3.30
18	36	26.95	5.07	Food & Drink	5	23	13.81	5.19
10	27	19.19	4.17	Sport & PA	4	15	10.38	3.02
8	32	21.95	6.38	Env & Climate	3	21	15.05	5.11
7	24	15.24	4.75	Economy & Money	2	20	11.00	4.69
70	142	105.04	19.58	Mean LAT	23	86	62.57	16.30

Table 8.23*Normality Tests in LAT for Gender*

9 th grade: Male						
English			Prompt	French		
<i>W</i>	<i>df</i>	<i>p</i>		<i>W</i>	<i>df</i>	<i>p</i>
.972	16	.872	Animals	.891	19	.033
.941	16	.358	Food & Drink	.945	19	.325
.865	16	.022	Sport & Physical Activities	.923	19	.128
.939	16	.340	Environment & Climate	.952	19	.421
.910	16	.118	Economy & Money	.911	19	.076
.892	16	.061	LAT Mean	.925	19	.138
9 th grade: Female						
English			Prompt	French		
<i>W</i>	<i>df</i>	<i>p</i>		<i>W</i>	<i>df</i>	<i>p</i>
.918	23	.061	Animals	.894	23	.019
.939	23	.170	Food & Drink	.928	22	.113
.966	23	.584	Sport & Physical Activities	.971	22	.745
.976	23	.823	Environment & Climate	.945	23	.233
.964	23	.550	Economy & Money	.953	23	.332
.945	23	.228	LAT Mean	.917	23	.058
10 th grade: Male						
English			Prompt	French		
<i>W</i>	<i>df</i>	<i>p</i>		<i>W</i>	<i>df</i>	<i>p</i>
.933	15	.307	Animals	.946	15	.463
.979	15	.964	Food & Drink	.944	15	.440
.981	15	.974	Sport & Physical Activities	.919	15	.184
.978	15	.956	Environment & Climate	.936	15	.335
.922	15	.209	Economy & Money	.962	15	.722
.976	15	.931	LAT Mean	.966	15	.789

Table 8.23 (continued)

10 th grade: Female						
English				French		
<i>W</i>	<i>df</i>	<i>p</i>	Prompt	<i>W</i>	<i>df</i>	<i>p</i>
.949	26	.214	Animals	.955	26	.301
.951	26	.250	Food & Drink	.979	26	.856
.964	26	.477	Sport & Physical Activities	.963	26	.454
.975	26	.752	Environment & Climate	.947	26	.196
.935	26	.102	Economy & Money	.972	26	.670
.965	26	.492	LAT Mean	.955	26	.304
11 th grade: Male						
English				French		
<i>W</i>	<i>df</i>	<i>p</i>	Prompt	<i>W</i>	<i>df</i>	<i>p</i>
.960	19	.564	Animals	.907	15	.123
.961	19	.592	Food & Drink	.918	15	.179
.931	19	.182	Sport & Physical Activities	.903	15	.105
.879	19	.021	Environment & Climate	.933	15	.304
.957	19	.508	Economy & Money	.941	15	.389
.925	19	.140	LAT Mean	.938	15	.361
11 th grade: Female						
English				French		
<i>W</i>	<i>df</i>	<i>p</i>	Prompt	<i>W</i>	<i>df</i>	<i>p</i>
.962	21	.549	Animals	.954	21	.407
.969	21	.713	Food & Drink	.958	21	.477
.972	21	.771	Sport & Physical Activities	.971	21	.766
.912	21	.061	Environment & Climate	.923	21	.101
.973	21	.803	Economy & Money	.976	21	.852
.976	21	.860	LAT Mean	.912	21	.061

In order to determine whether the above differences were statistically significant, paired samples t-tests and Wilcoxon signed-rank tests, in the case of the prompts *Animals* and *Sport & Physical Activities* for 9th grade male students, the prompt *Animals* for 9th grade female students and the prompt *Environment & Climate* for 11th grade male students, were carried out to compare the overall number of responses and the number of responses for each prompt produced by male participants in English and French and by and female participants in English and French in each of the three grades (Table 8.24).

Results revealed that both male and female participants in each grade produced a statistically significant higher number of words in English than in French overall. While statistically significant differences were also found in all five prompts by female participants in each grade and in almost all prompts by male participants, two exceptions were observed in the case of the latter: no statistically significant differences were found between English and French for the prompt *Economy & Money* in 9th grade or 10th grade male students. It should be recalled that while 9th grade students do not study economics, 10th grade students at the time of the first data collection had been studying economics through English for a period of four months. This does not appear to have been enough to make a difference between the two languages in the content-relevant prompt *Economy & Money*. However, at the time of the second data collection, 11th grade male students had been studying economics through English for an additional year. This amount of exposure may possibly have resulted in an improvement in their performance on the prompt *Economy & Money* for the English LAT, to the extent that a difference was found between the two languages in 11th grade but not the previous two grades. These results are presented visually in Figure 8.7.

Upon comparing the results of the English LAT and French LAT by male and female students, it is thus clear that both groups demonstrate a higher productive vocabulary in English as compared with French. These results are consistent with those found in RQ1, however, some exceptions are observed with regards to the content-relevant prompt *Economy & Money* for male participants. This will be discussed further in Section 8.4.1 with regard to CLIL and LAT.

Table 8.24*Differences in Lexical Availability in English and French by Gender*

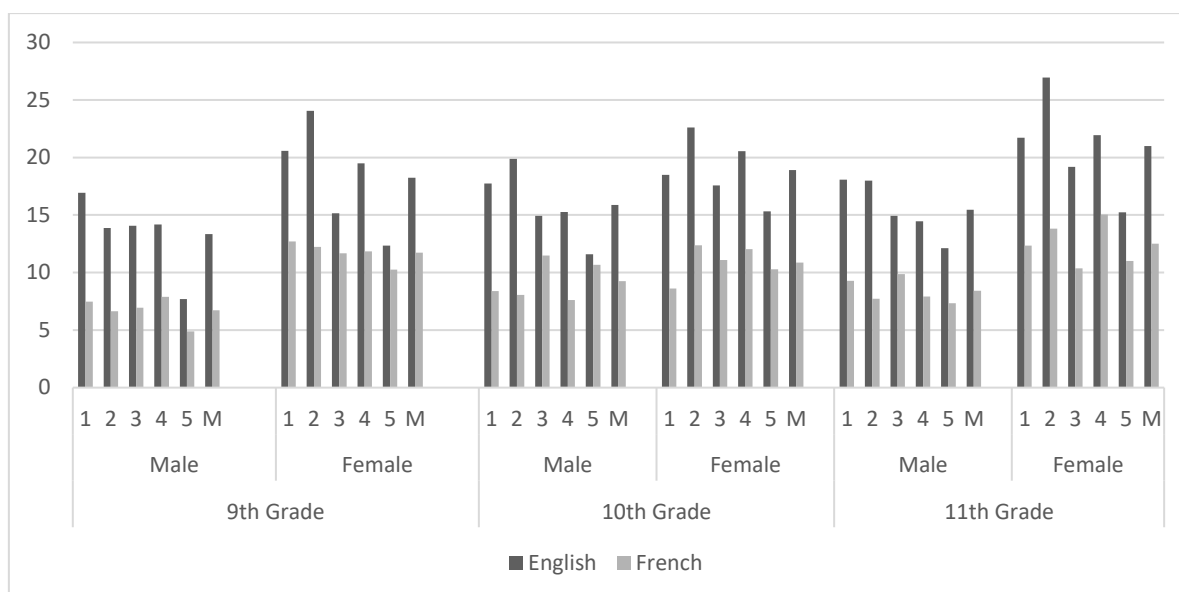
9 th grade: Male							
Prompt	<i>M</i>		<i>SD</i>		<i>t / z</i>	<i>df</i>	<i>p</i>
	English	French	English	French			
Animals*	16.94	7.47	8.31	3.53	-3.07		.002
Food & Drink	13.88	6.63	8.17	3.72	3.65	15	.002
Sport & Physical Activities*	14.06	6.95	7.10	3.70	-3.31		.001
Environment & Climate	14.19	7.88	8.47	5.35	3.25	15	.005
Economy & Money	7.69	4.88	6.61	3.24	2.05	15	.058
Mean LAT	13.35	6.72	6.90	3.00	4.27	15	.001
9 th grade: Female							
Prompt	<i>M</i>		<i>SD</i>		<i>t / z</i>	<i>df</i>	<i>p</i>
	English	French	English	French			
Animals*	20.57	12.70	5.86	5.98	-3.79		< .001
Food & Drink	24.05	12.23	6.40	4.73	8.10	21	< .001
Sport & Physical Activities	15.14	11.68	5.12	2.93	3.43	21	.002
Environment & Climate	19.48	11.83	6.29	4.68	6.34	22	< .001
Economy & Money	12.35	10.26	4.83	4.57	2.93	22	.008
Mean LAT	18.25	11.72	4.96	3.33	8.94	22	< .001
10 th grade: Male							
Prompt	<i>M</i>		<i>SD</i>		<i>t</i>	<i>df</i>	<i>p</i>
	English	French	English	French			
Animals	17.73	8.40	4.496	2.77	10.45	14	< .001
Food & Drink	19.87	8.07	4.734	4.31	10.36	14	< .001
Sport & Physical Activities	14.93	11.47	2.764	2.03	4.19	14	.001
Environment & Climate	15.27	7.60	6.19	3.66	6.46	14	< .001
Economy & Money	11.60	10.67	4.71	3.71	.788	14	.444
Mean LAT	15.88	9.24	3.41	2.58	11.80	14	< .001

Note. * = non-parametric test used, given that the data were not normally distributed.

Table 8.24 (continued)

10 th grade: Female							
Prompt	<i>M</i>		<i>SD</i>		<i>t</i>	<i>df</i>	<i>p</i>
	English	French	English	French			
Animals	18.50	8.62	5.08	3.87	9.92	25	< .001
Food & Drink	22.62	12.38	5.42	5.216	8.95	25	< .001
Sport & Physical Activities	17.58	11.08	5.13	2.85	6.98	25	< .001
Environment & Climate	20.54	12.04	4.79	4.60	8.40	25	< .001
Economy & Money	15.31	10.27	4.36	3.82	5.46	25	< .001
Mean LAT	18.90	10.87	3.96	3.19	11.47	25	< .001
11 th grade: Male							
Prompt	<i>M</i>		<i>SD</i>		<i>t / z</i>	<i>df</i>	<i>p</i>
	English	French	English	French			
Animals	18.07	9.27	6.00	3.93	7.50	14	< .001
Food & Drink	18.00	7.73	5.63	4.20	11.65	14	< .001
Sport & Physical Activities	14.93	9.87	4.55	3.33	6.09	14	< .001
Environment & Climate*	14.47	7.93	6.52	4.57	-3.18		.001
Economy & Money	12.13	7.33	5.20	3.374	3.77	14	.002
Mean LAT	15.45	8.42	4.35	3.14	12.95	14	< .001
11 th grade: Female							
Prompt	<i>M</i>		<i>SD</i>		<i>t</i>	<i>df</i>	<i>p</i>
	English	French	English	French			
Animals	21.71	12.33	5.36	3.30	8.29	20	< .001
Food & Drink	26.95	13.81	5.07	5.19	10.51	20	< .001
Sport & Physical Activities	19.19	10.38	4.17	3.02	8.31	20	< .001
Environment & Climate	21.95	15.05	6.38	5.11	5.85	20	< .001
Economy & Money	15.24	11.00	4.75	4.69	3.99	20	.001
Mean LAT	21.00	12.51	3.91	3.26	9.93	20	< .001

Note. * = non-parametric test used, given that the data were not normally distributed.

Figure 8.7*Differences in LAT within Gender*

Note. 1 = Animals, 2 = Food & Drink, 3 = Sport and Physical Activities, 4 = Environment & Climate, 5 = Economy & Money, M = Mean LAT.

In order to determine whether there were cross-sectional and longitudinal differences across gender in the English and French LATs, two-way mixed ANOVAS were carried out comparing participants in 9th and 10th grade, including one within-subjects factor (language) and one between-subjects factor (time), and two-way repeated ANOVAS were carried out comparing participants in 10th and 11th grade, with two within-subjects factors (time and language). This was done first for males and then for females in each case.

Firstly, regarding cross-sectional differences in male students, two-way mixed ANOVAS were carried out on the LATs in each language (Table 8.25). Results indicated that there was no significant main effect of time on the overall LATs ($F(1, 29) = 3.64, p = .066, \eta_p^2 = .112$), where means increased from 9th grade ($M = 10.03$) to 10th grade ($M = 12.56$). There was, however, a statistically significant main effect of language ($F(1, 29) = 61.45, p < .001, \eta_p^2 = .679$), with mean scores for the English LAT ($M = 14.61$) much higher than for the French LAT ($M = 7.98$). There was no statistically significant interaction between time and language for male students ($F(1, 29) = .000, p = .993, \eta_p^2 = .000$); male participants' score on the English LAT increased from 9th grade ($M = 13.35$) to 10th grade ($M = 15.88$), and their score on the French LAT increased from 9th grade ($M = 6.72$) to 10th grade ($M = 9.24$). Regarding the individual prompts, results indicated that,

while means were higher for all prompts in 10th grade participants, there was no significant main effect of time on the prompts *Animals*, *Sport & Physical Activity* and *Environment & Climate*, but there was a significant main effect of time on the prompts *Food & Drink* and *Economy & Money*. A statistically significant main effect of language was found in all prompts, with means for English being a great deal higher than French across these grades. There was no statistically significant interaction between time and language for male students for any of the individual prompts, suggesting that performance in one grade or the other was not dependent on language. In order to determine whether there was an effect of time or an interaction between time and language on male participants' language level, two-way mixed ANOVAS were again carried out. The effect of language was not considered here, given that the English and French C-tests are not directly comparable, as previously indicated. Results indicated that there was a significant main effect of time on the C-tests ($F(1, 29) = 11.00, p = .002, \eta_p^2 = .262$), with means increasing from 9th grade ($M = 24.72$) to 10th grade ($M = 34.36$). There was no statistically significant interaction between time and language for male students ($F(1, 29) = 1.01, p = .322, \eta_p^2 = .031$); male participants' improved their score on the English C-test from 9th grade ($M = 22.27$) to 10th grade ($M = 33.80$), and their score on the French C-test from 9th grade ($M = 27.16$) to 10th grade ($M = 34.93$).

Secondly, regarding cross-sectional differences in female students, two-way mixed ANOVAS were carried out on the LATs in each language (Table 8.26). Results indicated that there was no significant main effect of time on the overall LATs ($F(1, 47) = .010, p = .922, \eta_p^2 = .000$), where means decreased slightly from 9th grade ($M = 14.99$) to 10th grade ($M = 14.89$). There was, however, a statistically significant main effect of language ($F(1, 47) = 206.68, p < .001, \eta_p^2 = .815$), with mean scores for the English LAT ($M = 18.58$) much higher than for the French LAT ($M = 11.30$). This was also found in the case of male students. There was no statistically significant interaction between time and language for female students ($F(1, 47) = 2.21, p = .144, \eta_p^2 = .045$); female participants' score on the English LAT increased from 9th grade ($M = 18.25$) to 10th grade ($M = 18.91$), whereas their score on the French LAT decreased from 9th grade ($M = 11.73$) to 10th grade ($M = 10.88$). Regarding the individual prompts, results indicated that there was a significant main effect of time only on the prompt *Animals*, where 9th grade students had higher means than 10th grade students in the two LATs. Means were also higher in 9th grade for the prompt *Food & Drink*, while they were higher for all content-relevant prompts, *Sport & Physical Activity*, *Environment & Climate* and *Economy & Money*, in

10th grade. A statistically significant main effect of language was again found in all prompts, with means for English being a great deal higher than French across these grades. While there was no statistically significant interaction between time and language for male students for any of the individual prompts, in the case of female students, a significant interaction was found for the two English-related prompts: *Sport & Physical Activity* ($F(1, 47) = 4.93, p = .0312, \eta_p^2 = .097$) and *Economy & Money* ($F(1, 47) = 6.17, p = .017, \eta_p^2 = .116$). In the prompt *Sport & Physical Activity*, female participants slightly increased their score in English from 9th grade ($M = 15.14$) to 10th grade ($M = 17.58$), and slightly decreased their score in French from 9th grade ($M = 11.68$) to 10th grade ($M = 11.08$). In the prompt *Economy & Money*, female participants slightly increased their score in English from 9th grade ($M = 12.35$) to 10th grade ($M = 15.31$), while their score in French from 9th grade ($M = 10.26$) to 10th grade ($M = 10.27$) remained largely the same. This finding is likely related to the increased exposure to content-related vocabulary in the participants' English CLIL classes. In order to determine whether there was an effect of time or an interaction between time and language on female participants' language level, two-way mixed ANOVAS were again carried out. The effect of language was again not considered here, given that the English and French C-tests are not directly comparable, as previously indicated. Results revealed that, unlike in the case of male participants, there was no significant main effect of time on the C-test ($F(1, 47) = .036, p = .851, \eta_p^2 = .001$), with only minor improvements from 9th grade ($M = 34.33$) to 10th grade ($M = 34.96$). There was also no statistically significant interaction between time and language for male students ($F(1, 47) = .746, p = .392, \eta_p^2 = .016$); female participants slightly decreased their score on the English C-test from 9th grade ($M = 34.52$) to 10th grade ($M = 34.19$), and slightly increased their score on the French C-test from 9th grade ($M = 34.13$) to 10th grade ($M = 35.73$).

Table 8.25*Cross-Sectional Differences for Male Participants in the English and French LATs*

Time							
	<i>M</i>		<i>F</i>	<i>p</i>	η_p^2		
	9 th Grade	10 th Grade					
1	11.96	13.06	.587	.450	.020		
2	10.25	13.96	5.27	.029	.154		
3	10.65	13.20	3.51	.071	.108		
4	11.03	11.43	.045	.834	.002		
5	6.28	11.13	11.18	.002	.278		
Mean	10.03	12.56	3.64	.066	.112		
Language							
	<i>M</i>		<i>F</i>	<i>P</i>	η_p^2		
	English	French					
1	17.33	7.70	59.45	< .001	.672		
2	16.87	7.34	67.03	< .001	.698		
3	14.49	9.35	35.71	< .001	.552		
4	14.72	7.73	36.59	< .001	.558		
5	9.64	7.77	4.22	.049	.127		
Mean	14.61	7.98	61.45	< .001	.679		
Time*Language							
	<i>M</i>				<i>F</i>	<i>p</i>	η_p^2
	9 th Grade		10 th Grade				
	English	French	English	French			
1	16.93	7.00	17.73	8.40	.058	.811	.002
2	13.87	6.62	19.86	8.06	3.82	.060	.117
3	14.06	7.25	14.93	11.46	3.78	.061	.115
4	14.18	7.87	15.26	7.60	.343	.562	.012
5	7.68	4.87	11.60	10.66	1.06	.311	.035
Mean	13.35	6.72	15.88	9.24	.000	.993	.000

Note. Prompt 1 = *Animals*, Prompt 2 = *Food & Drink*, Prompt 3 = *Sport and Physical Activities*, Prompt 4 = *Environment & Climate*, Prompt 5 = *Economy & Money*.

Table 8.26*Cross-Sectional Differences for Female Participants in the English and French LATs*

Time							
	<i>M</i>		<i>F</i>	<i>p</i>	η_p^2		
	9 th Grade	10 th Grade					
1	16.63	13.56	5.64	.022	.107		
2	18.14	17.50	.242	.625	.005		
3	13.41	14.33	.855	.360	.018		
4	15.65	16.29	.264	.610	.006		
5	11.30	12.79	1.79	.187	.037		
Mean	14.99	14.89	.010	.922	.000		
Language							
	<i>M</i>		<i>F</i>	<i>p</i>	η_p^2		
	English	French					
1	19.53	10.66	138.73	< .001	.747		
2	23.33	12.31	145.63	< .001	.760		
3	16.36	11.38	52.74	< .001	.534		
4	20.01	11.93	106.91	< .001	.695		
5	13.83	10.27	36.00	< .001	.434		
Mean	18.58	11.30	206.68	< .001	.815		
Time*Language							
	<i>M</i>				<i>F</i>	<i>p</i>	η_p^2
	9 th Grade		10 th Grade				
	English	French	English	French			
1	20.57	12.70	18.50	8.62	1.78	.188	.037
2	24.05	12.23	22.62	12.39	.755	.389	.016
3	15.14	11.68	17.58	11.08	4.93	.031	.097
4	19.48	11.83	20.54	12.04	.295	.590	.006
5	12.35	10.26	15.31	10.27	6.17	.017	.116
Mean	18.25	11.73	18.91	10.88	2.21	.144	.045

Note. Prompt 1 = *Animals*, Prompt 2 = *Food & Drink*, Prompt 3 = *Sport and Physical Activities*, Prompt 4 = *Environment & Climate*, Prompt 5 = *Economy & Money*.

Thirdly, regarding longitudinal differences in male students, two-way repeated ANOVAs were carried out on the LATs in each language (Table 8.27). Results revealed no significant main effect of time on the overall LATs ($F(1, 11) = .156, p = .700, \eta_p^2 = .014$), where means actually decreased from 10th grade ($M = 12.37$) to 11th grade ($M = 12.08$). There was, however, a statistically significant main effect of language ($F(1, 11) = 481.02, p < .001, \eta_p^2 = .978$), with mean scores for the English LAT ($M = 15.64$) much higher than for the French LAT ($M = 8.81$). There was no statistically significant interaction between time and language for male students ($F(1, 11) = .069, p = .797, \eta_p^2 = .006$); male participants' score on the English LAT decreased slightly from 10th grade ($M = 15.75$) to 11th grade ($M = 15.53$), and their score on the French LAT decreased slightly from 10th grade ($M = 9.00$) to 11th grade ($M = 8.63$). Regarding the individual prompts, results indicated that there was no significant main effect of time on any of the five prompts, where means for *Animals* and *Environment & Climate* were slightly higher in 11th grade whereas means for *Food & Drink*, *Sport & Physical Activity* and *Economy & Money* were slightly higher in 10th grade. A statistically significant main effect of language was found in all prompts, with means for English being a great deal higher than French across these grades. There was no statistically significant interaction between time and language for male students for the majority of the prompts, with just one exception: *Economy & Money* ($F(1, 11) = 10.82, p = .007, \eta_p^2 = .496$). Result showed that male participants' score on this prompt in English increased from 10th grade ($M = 11.50$) to 11th grade ($M = 12.25$) while their score on the French LAT decreased quite a bit from 10th grade ($M = 10.58$) to 11th grade ($M = 7.08$), and that while results in each language were relatively similar in 10th grade, a much greater difference was seen in 11th grade. In order to determine whether there was an effect of time or interaction between time and language on male participants' language level, two-way repeated ANOVAs were again carried out. The effect of language was again not considered here, given that the English and French C-tests are not directly comparable, as previously stated. Results indicated that there was no significant main effect of time on the C-tests ($F(1, 11) = 2.94, p = .114, \eta_p^2 = .211$), though means increased from 10th grade ($M = 31.37$) to 11th grade ($M = 35.70$). There was no statistically significant interaction between time and language for male students ($F(1, 11) = .178, p = .681, \eta_p^2 = .016$); male participants improved their score on the English C-test from 10th grade ($M = 29.08$) to 11th grade ($M = 33.66$), and improved their score on the French C-test from 10th grade ($M = 32.25$) to 11th grade ($M = 39.16$).

Finally, regarding longitudinal differences in female students, two-way repeated

ANOVAs were carried out on the LATs in each language (Table 8.28). Results indicated that there was a statistically significant main effect of time on the overall LATs ($F(1, 11) = 25.45, p < .001, \eta_p^2 = .600$), where means increased from 10th grade ($M = 15.77$) to 11th grade ($M = 17.12$). There was also a statistically significant main effect of language ($F(1, 11) = 92.04, p < .001, \eta_p^2 = .844$), with mean scores for the English LAT ($M = 20.46$) much higher than for the French LAT ($M = 12.44$). There was no statistically significant interaction between time and language for female students ($F(1, 17) = .090, p = .767, \eta_p^2 = .005$); female participants' score on the English LAT increased slightly from 10th grade ($M = 19.83$) to 11th grade ($M = 21.08$), and their score on the French LAT increased slightly from 10th grade ($M = 11.72$) to 11th grade ($M = 13.16$). Regarding the individual prompts, results indicated, unlike the case of male participants, time appeared to play a greater role in some prompts. While no significant main effect of time was found for the two English-related prompts, *Sport & Physical Activity* and *Economy & Money*, a significant main effect of time was found for *Animals*, *Food & Drink* and *Environment & Climate*, which saw a marked improvement from 10th to 11th grade. As in the case of male students, a statistically significant main effect of language was found in all prompts, with means for English being a great deal higher than French across these grades. There was no statistically significant interaction between time and language for male students for the majority of the prompts, with just one exception: *Food & Drink* ($F(1, 17) = 4.70, p = .045, \eta_p^2 = .217$). Results showed that female participants' score on this prompt in English increased greatly from 10th grade ($M = 23.55$) to 11th grade ($M = 27.05$) while their score on the French LAT increased just slightly from 10th grade ($M = 13.44$) to 11th grade ($M = 14.88$). A much larger difference between means in each language was thus observed in 11th grade than in 10th grade. In order to determine whether there was an effect of time or interaction between time and language on female participants' language level, two-way repeated ANOVAs were again carried out. The effect of language was again not considered here, given that the English and French C-tests are not directly comparable, as previously stated. Results indicated that there was no significant main effect of time on the C-tests ($F(1, 17) = 4.46, p = .050, \eta_p^2 = .208$), though means increased from 10th grade ($M = 38.22$) to 11th grade ($M = 43.19$). There was no statistically significant interaction between time and language for female students ($F(1, 17) = 1.67, p = .213, \eta_p^2 = .090$); female participants' improved their score on the English C-test from 10th grade ($M = 38.94$) to 11th grade ($M = 41.50$), and their score on the French C-test from 10th grade ($M = 37.50$) to 11th grade ($M = 44.88$).

Table 8.27*Longitudinal Differences for Male Participants in the English and French LATs*

Time							
	<i>M</i>				<i>F</i>	<i>p</i>	η_p^2
	10 th Grade		11 th Grade				
1	12.58		13.91		3.061	.108	.218
2	13.83		12.79		.986	.342	.082
3	13.04		12.62		.189	.672	.017
4	11.37		11.41		.002	.969	.000
5	11.04		9.66		1.43	.257	.115
Mean	12.37		12.08		.156	.700	.014
Language							
	<i>M</i>				<i>F</i>	<i>p</i>	η_p^2
	English		French				
1	17.62		8.87		165.00	< .001	.938
2	19.04		7.58		222.96	< .001	.953
3	15.20		10.45		39.31	< .001	.781
4	14.45		8.33		40.90	< .001	.788
5	11.87		8.33		6.52	.027	.372
Mean	15.64		8.81		481.02	< .001	.978
Time*Language							
	<i>M</i>				<i>F</i>	<i>p</i>	η_p^2
	10 th Grade		11 th Grade				
	English	French	English	French			
1	17.08	8.08	18.16	9.66	.133	.723	.012
2	20.08	7.58	18.00	7.58	2.37	.151	.178
3	15.25	10.83	15.16	10.08	.662	.433	.057
4	14.83	7.91	14.08	8.75	1.29	.279	.106
5	11.50	10.58	12.25	7.08	10.82	.007	.496
Mean	15.75	9.00	15.53	8.63	.069	.797	.006

Note. Prompt 1 = *Animals*, Prompt 2 = *Food & Drink*, Prompt 3 = *Sport and Physical Activities*, Prompt 4 = *Environment & Climate*, Prompt 5 = *Economy & Money*.

Table 8.28*Longitudinal Differences for Female Participants in the English and French LATs*

Time							
	<i>M</i>				<i>F</i>	<i>p</i>	η_p^2
	10 th Grade		11 th Grade				
1	14.19		17.13		28.54	< .001	.627
2	18.50		20.97		9.75	.006	.365
3	15.44		14.72		1.53	.232	.083
4	17.13		19.44		10.42	.005	.380
5	13.61		13.36		.175	.681	.010
Mean	15.77		17.12		25.45	< .001	.600
Language							
	<i>M</i>				<i>F</i>	<i>p</i>	η_p^2
	English		French				
1	20.44		10.88		89.14	< .001	.840
2	25.30		14.16		68.85	< .001	.802
3	18.77		11.38		52.24	< .001	.754
4	22.13		14.44		49.89	< .001	.746
5	15.63		11.33		29.45	< .001	.634
Mean	20.46		12.44		92.04	< .001	.844
Time*Language							
	<i>M</i>				<i>F</i>	<i>p</i>	η_p^2
	10 th Grade		11 th Grade				
	English	French	English	French			
1	19.16	9.22	21.72	12.55	.278	.605	.016
2	23.55	13.44	27.05	14.88	4.703	.045	.217
3	18.94	11.94	18.61	10.83	.989	.334	.055
4	21.44	12.83	22.83	16.05	1.910	.185	.101
5	16.05	11.16	15.22	11.50	1.000	.331	.056
Mean	19.83	11.72	21.08	13.16	.090	.767	.005

Note. Prompt 1 = *Animals*, Prompt 2 = *Food & Drink*, Prompt 3 = *Sport and Physical Activities*, Prompt 4 = *Environment & Climate*, Prompt 5 = *Economy & Money*.

In summary, results of ANOVAs revealed some interesting gender-based differences, in particular concerning time and the interaction between time and language. Firstly, regarding time, the cross-sectional analysis indicated that while for male participants there was significant main effect of time on the C-tests, the prompt *Food & Drink* and the prompt *Economy & Money*, with higher means in 10th grade, for female participants there was significant main effect of time only on the prompt *Animals*, with higher means in 9th grade. The longitudinal analysis also showed clear gender-based differences. While for male participants there was no significant main effect of time on the C-tests, the overall LAT or any of the five individual prompts, for female participants there was a significant main effect of time on the overall LAT and the individual prompts *Animals*, *Food & Drink*, and *Environment & Climate*, with statistically significant higher means in 11th grade. This suggests that female students improved to a much greater degree than male students both in their overall LA and specifically within these lexical domains. Secondly, regarding language, both the cross-sectional and longitudinal analysis indicated that for both male and female participants, there was a significant main effect of language on the overall LAT and each of the five individual prompts. In all cases, means in English were higher than in French. This highlights the higher lexical proficiency that students have in English as compared to French across the grades at hand. Finally, regarding the interaction between time and language, the cross-sectional analysis indicated that for male participants there were no statistically significant interactions between time and language in the C-tests, the overall LAT or any of the five individual prompts. Meanwhile, for female students, statistically significant interactions were found in the individual prompts *Sport & Physical Activity* and *Economy & Money*. From 9th to 10th grade, means for *Sport & Physical Activity* increased in English and decreased in French while for *Economy & Money* they increased in English and remained largely the same in French. These differences may likely be attributable to the increased exposure to content-related vocabulary in the participants' English CLIL classes, physical education and economics. The longitudinal analysis indicated that there was a statistically significant interaction between time and language for male participants in the prompt *Economy & Money* and for female participants in the prompt *Food & Drink*. Male participants' score in the prompt *Economy & Money* increased in English from 10th to 11th grade, perhaps due to the exposure that was received in their English CLIL class, economics, while it decreased in French. Female participants' score in the prompt *Food & Drink* increased greatly in English and only slightly in French from 10th to 11th grade,

and a much larger difference was observed between means in each language in 11th grade than in 10th grade, suggesting a significant improvement in English in this lexical domain as compared to French. Table 8.29 provides a summary of these differences, while Figures 8.8-8.11 provide a visual representation of the interactions.

Table 8.29

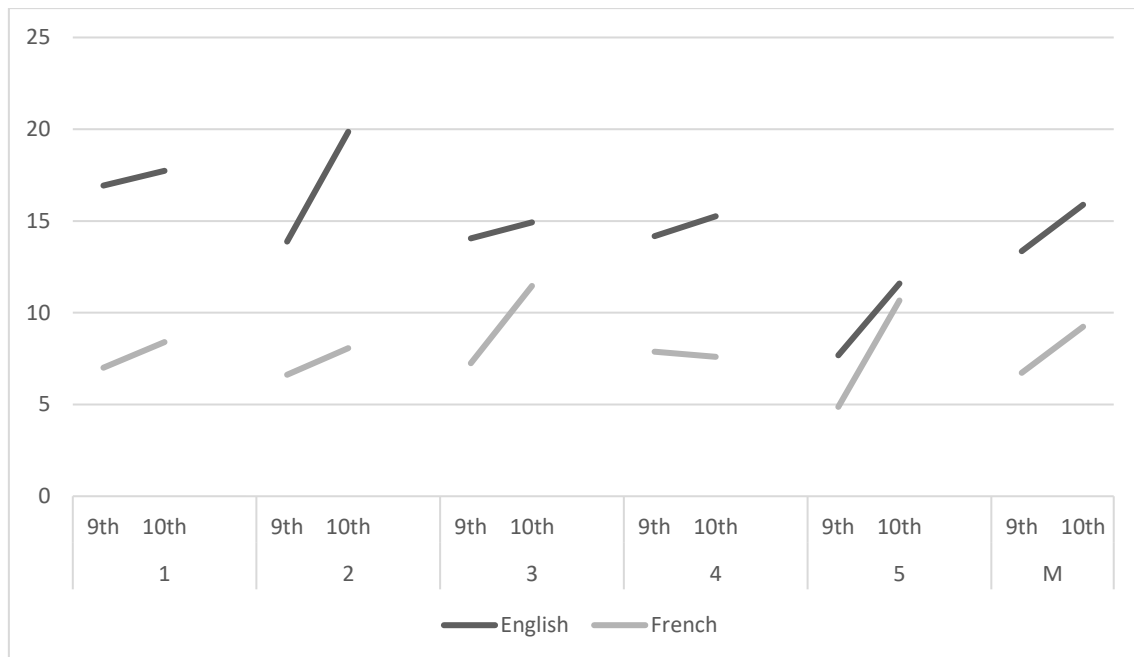
Summary of Cross-Sectional and Longitudinal Differences in English and French C-tests and LATs by Gender

Time				
	9 th Grade → 10 th Grade		10 th → 11 th Grade	
	Male	Female	Male	Female
C-test	✓			✓
1		✓		✓
2	✓			✓
3				
4				✓
5	✓			
M				
Language				
	9 th Grade → 10 th Grade		10 th → 11 th Grade	
	Male	Female	Male	Female
1	✓	✓	✓	✓
2	✓	✓	✓	✓
3	✓	✓	✓	✓
4	✓	✓	✓	✓
5	✓	✓	✓	✓
M	✓	✓	✓	✓
Time*Language				
	9 th Grade → 10 th Grade		10 th → 11 th Grade	
	Male	Female	Male	Female
C-test				
1				
2				✓
3		✓		
4				
5		✓	✓	
M				

Note. ✓ = statistically significant differences, 1 = *Animals*, 2 = *Food & Drink*, 3 = *Sport and Physical Activities*, 4 = *Environment & Climate*, 5 = *Economy & Money*, M = Mean LAT.

Figure 8.8

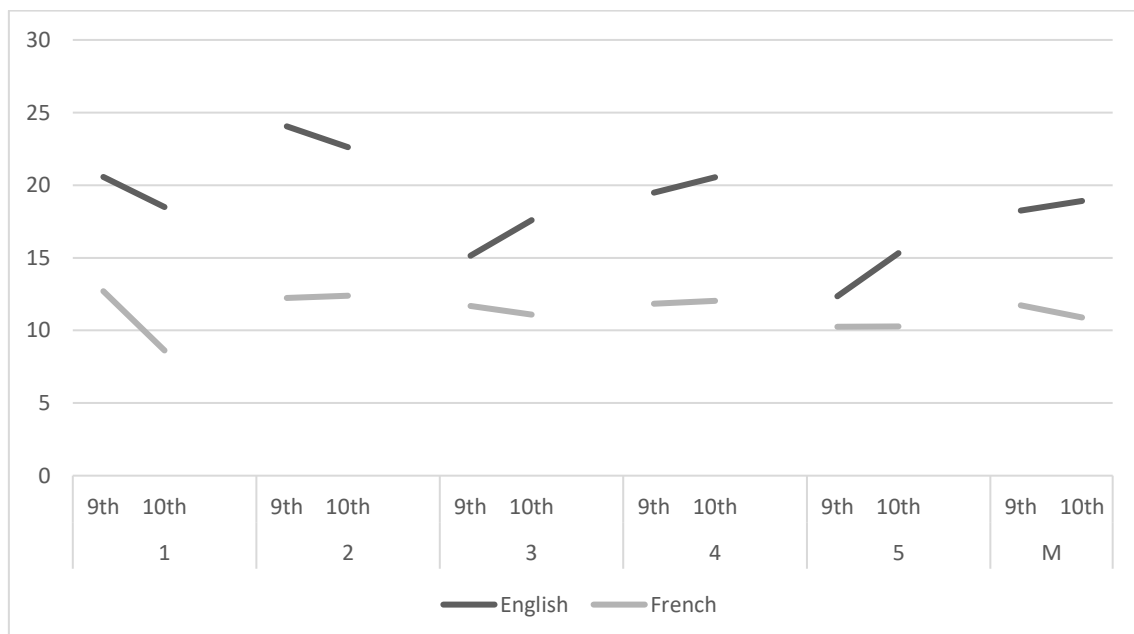
Cross-Sectional Differences for Male Participants in the English and French LATs



Note. 1 = Animals, 2 = Food & Drink, 3 = Sport and Physical Activities, 4 = Environment & Climate, 5 = Economy & Money, M = Mean LAT.

Figure 8.9

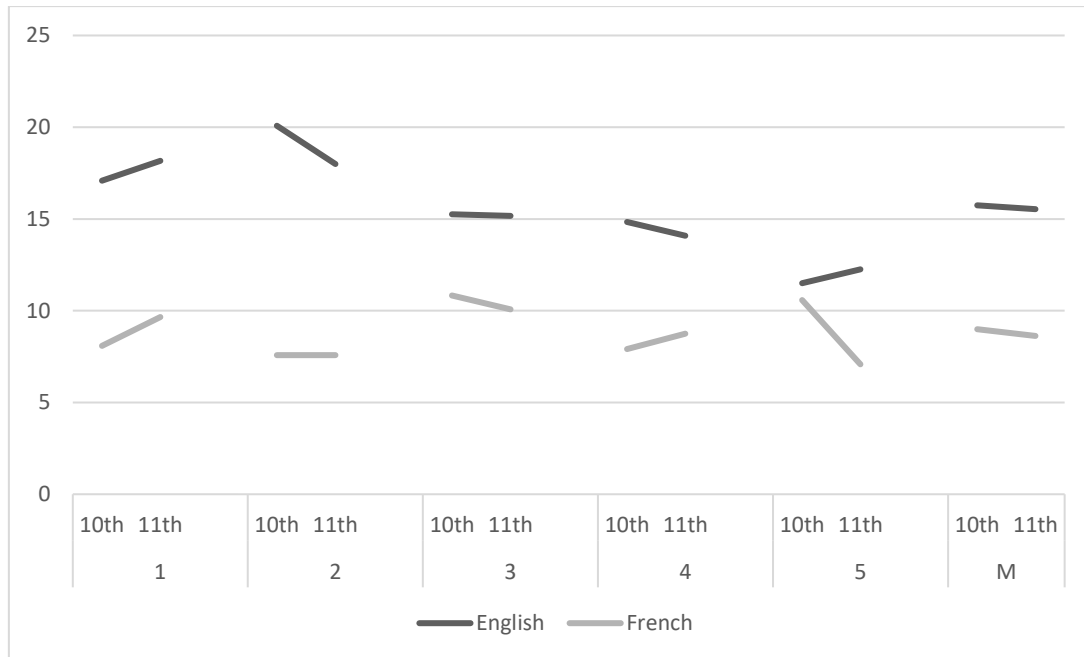
Cross-Sectional Differences for Female Participants in the English and French LATs



Note. 1 = Animals, 2 = Food & Drink, 3 = Sport and Physical Activities, 4 = Environment & Climate, 5 = Economy & Money, M = Mean LAT.

Figure 8.10

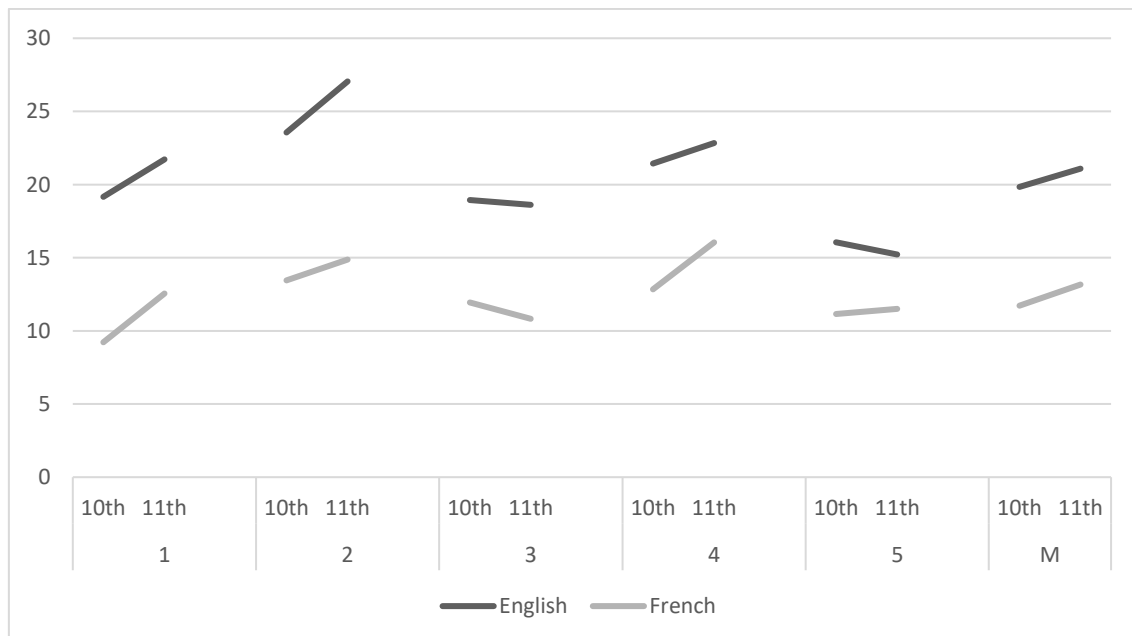
Longitudinal Differences for Male Participants in the English and French LATs



Note. 1 = Animals, 2 = Food & Drink, 3 = Sport and Physical Activities, 4 = Environment & Climate, 5 = Economy & Money, M = Mean LAT.

Figure 8.11

Longitudinal Differences for Female Participants in the English and French LATs



Note. 1 = Animals, 2 = Food & Drink, 3 = Sport and Physical Activities, 4 = Environment & Climate, 5 = Economy & Money, M = Mean LAT.

8.3.1.2. Quantitative Differences in Gender and LA within Languages. Research question 3.1.2 asked whether there were quantitative differences in the words retrieved by male participants as compared to female participants in English and French in each grade at each testing period. Independent samples t-tests and Mann-Whitney U tests, in the case of the prompts *Animals* in English for 9th grade students, *Animals* and *Sport & Physical Activities* in French for 9th grade students, and *Environment & Climate* in English for 11th grade students, were thus conducted to determine if there were differences between the male and female students in each grade with regard to LA, first in English and then in French.

With regard to English, the descriptive statistics for the overall LAT indicated that at all levels the female groups produced a higher number of words than the male groups, both in the overall LAT and in the five individual prompts (Table 8.30). As outlined in the previous section, data were normally distributed in all cases with the exception of the prompt *Animals* in English for 9th grade male students and *Environment & Climate* in English for 11th grade male students. Independent samples t-tests and Mann-Whitney U tests, in the case of the prompt *Animals* in 9th grade and *Environment & Climate* in 11th grade, were carried out to determine whether these differences were statistically significant (Table 8.31). Results showed that there were statistically significant differences in the overall English LAT between the male group ($M = 13.35$, $SD = 6.90$) and the female group ($M = 18.25$, $SD = 4.96$) in 9th grade ($t(37) = -2.58$, $p = .014$), between the male group ($M = 15.88$, $SD = 3.41$) and the female group ($M = 18.90$, $SD = 3.96$) in 10th grade ($t(39) = -2.47$, $p = .018$), and between the male group ($M = 15.64$, $SD = 4.16$) and the female group ($M = 21.00$, $SD = 3.91$) in 11th grade ($t(38) = -4.20$, $p < .001$). In all cases, female participants produced a higher number of tokens on average than male participants. Regarding the individual prompts, results of the independent samples t-tests and Mann-Whitney U tests revealed a number of differences between genders in terms of the individual prompts in English at each grade. In 9th grade, statistically significant differences were found in three out of the five prompts: *Food & Drink*, *Environment & Climate* and *Economy & Money*. However, no statistically significant differences were found for the prompts *Animals* and *Sport & Physical Activities*. In 10th grade, statistically significant differences were found in three out of the five prompts, namely, those which were related to their content classes: *Sport & Physical Activities*, *Environment & Climate* and *Economy & Money*. However, no statistically significant differences were found for the general prompts *Animals* or *Food & Drink*.

Table 8.30*Descriptive Statistics for Lexical Availability by Gender in English*

Male (<i>n</i> = 16)				9 th Grade	Female (<i>n</i> = 23)			
Min	Max	<i>M</i>	<i>SD</i>	Prompt	Min	Max	<i>M</i>	<i>SD</i>
1	31	16.94	8.31	Animals	13	32	20.57	5.86
0	35	13.88	8.17	Food & Drink	8	32	23.87	6.31
3	29	14.06	7.10	Sport & PA	4	24	15.00	5.04
4	33	14.19	8.47	Env & Climate	9	34	19.48	6.29
0	24	7.69	6.61	Eco & Money	1	20	12.35	4.83
21	151	66.75	34.52	Mean LAT	38	129	91.26	24.80
Male (<i>n</i> = 15)				10 th Grade	Female (<i>n</i> = 26)			
Min	Max	<i>M</i>	<i>SD</i>	Prompt	Min	Max	<i>M</i>	<i>SD</i>
12	28	17.73	4.49	Animals	10	28	18.50	5.08
12	29	19.87	4.73	Food & Drink	15	35	22.62	5.42
10	20	14.93	2.76	Sport & PA	9	27	17.58	5.13
2	26	15.27	6.19	Env & Climate	12	31	20.54	4.79
4	18	11.60	4.71	Eco & Money	10	26	15.31	4.36
46	107	79.40	17.06	Mean LAT	62	146	94.53	19.81
Male (<i>n</i> = 19)				11 th Grade	Female (<i>n</i> = 21)			
Min	Max	<i>M</i>	<i>SD</i>	Prompt	Min	Max	<i>M</i>	<i>SD</i>
7	29	18.05	5.47	Animals	12	30	21.71	5.36
8	28	18.11	6.42	Food & Drink	18	36	26.95	5.07
9	21	15.32	4.37	Sport & PA	10	27	19.19	4.17
4	26	14.47	6.52	Env & Climate	8	32	21.95	6.38
2	20	12.26	4.74	Eco & Money	7	24	15.24	4.75
48	112	78.21	20.81	Mean LAT	70	142	105.04	19.58

Table 8.31*Differences in Lexical Availability in English by Gender*

9 th Grade							
Prompt	<i>M</i>		<i>SD</i>		<i>t</i>	<i>df</i>	<i>p</i>
	Male	Female	Male	Female			
Animals*	16.94	20.57	8.31	5.86	-1.47		.141
Food & Drink	13.88	23.87	8.17	6.31	-4.30	37	<.001
Sport & Physical Activities	14.06	15.00	7.10	5.04	-.48	37	.632
Environment & Climate	14.19	19.48	8.47	6.29	-2.24	37	.031
Economy & Money	7.69	12.35	6.61	4.83	-2.54	37	.015
Mean LAT	13.35	18.25	6.90	4.96	-2.58	37	.014
10 th Grade							
Prompt	<i>M</i>		<i>SD</i>		<i>t/z</i>	<i>df</i>	<i>p</i>
	Male	Female	Male	Female			
Animals	17.73	18.50	4.49	5.08	-.484	39	.631
Food & Drink	19.87	22.62	4.73	5.42	-1.63	39	.110
Sport & Physical Activities	14.93	17.58	2.76	5.13	-2.14	38.89	0.39
Environment & Climate*	15.27	20.54	6.19	4.79	-3.04	39	.004
Economy & Money	11.60	15.31	4.71	4.36	-2.54	39	.015
Mean LAT	15.88	18.90	3.41	3.96	.423	39	.018
11 th Grade							
Prompt	<i>M</i>		<i>SD</i>		<i>t</i>	<i>df</i>	<i>p</i>
	Male	Female	Male	Female			
Animals	18.05	21.71	5.47	5.36	-2.13	38	.039
Food & Drink	18.11	26.95	6.42	5.07	-4.85	38	<.001
Sport & Physical Activities	15.32	19.19	4.37	4.17	-2.86	38	.007
Environment & Climate*	14.47	21.95	6.52	6.38	-3.66		.001
Economy & Money	12.26	15.24	4.74	4.75	-1.97	38	.055
Mean LAT	15.64	21.00	4.16	3.91	-4.20	38	<.001

Note. * = non-parametric test used, given that the data were not normally distributed.

Finally, in 11th grade, statistically significant differences were found in all prompts with the exception of one: *Economy & Money*. This finding is incredibly important, given the fact that economics is the one subject in which the language of instruction differs between male and female students in 11th grade: while male students continue to study economics through English in 11th grade, female students do so through Spanish. It is thus possible that continuing to study this subject via the TL has allowed the male students to effectively bridge the gap between them and their female peers in the prompt *Economy & Money*. This will be discussed further in Section 8.4.1 with regard to CLIL and LAT.

The above results show that the female group produced a higher number of tokens than the male group in the English LAT overall, a finding which was statistically significant at all levels. In addition, with regard to the individual prompts, *Environment & Climate* is the only prompt in which there were statistically significant differences at all levels. This indicates that there is potentially a female advantage with regard to this lexical domain. Three prompts revealed statistically significant differences at two levels, namely *Food & Drink* (9th and 11th grade), *Sport & Physical Activities* (10th and 11th grade) and *Economy & Money* (9th and 10th grade). Finally, statistically significant differences were found for the prompt *Animals* only in 11th grade.

In order to determine if these findings were consistent with the results of the language level C-test, the normality of the results in English were first assessed for each subgroup of gender in all grades (Table 8.32).

Table 8.32

Normality Tests for the English C-test for Gender

	9 th Grade			10 th Grade			11 th Grade		
	<i>W</i>	<i>df</i>	<i>p</i>	<i>W</i>	<i>df</i>	<i>p</i>	<i>W</i>	<i>df</i>	<i>p</i>
Male	.938	18	.271	.961	15	.705	.933	19	.198
Female	.960	23	.462	.954	26	.284	.979	21	.903

As shown, data were normally distributed at all grades, and so parametric tests were used. Independent samples t-tests were thus carried out to compare the English language level of male and female students in each grade. Results indicated that there was a statistically significant difference between the male group ($M = 22.28$, $SD = 8.40$) and the female group ($M = 34.52$, $SD = 12.47$) in 9th grade ($t(39) = -3.57$, $p = .001$), between the male group ($M = 28.20$, $SD = 12.97$) and the female group ($M = 37.42$, $SD = 9.93$) in 10th grade ($t(39) = -2.55$, $p = .015$), and between the male group ($M = 32.42$, $SD = 12.30$) and the

female group ($M = 42.19$, $SD = 11.04$) in 11th grade ($t(38) = -2.64$, $p = .012$). Again, in all cases, the female participants had a higher score on the English C-test than their male peers. It is thus highly likely that, in terms of gender, the overall results of the English LAT are related to English language proficiency.

With regard to French, the descriptive statistics for the overall LAT indicated that at all levels the female groups again produced a higher number of words on average than the male groups, and in almost all grades, with just two exceptions. In 10th grade, male students produced a higher number of words than female students in the categories *Sport & Physical Activities* and *Economy & Money* (Table 8.33). As outlined in the previous section, data were normally distributed in all cases with the exception of the prompt *Sport & Physical Activities* for 9th grade male students and *Animals* in 9th grade female students. Independent samples t-tests and Mann-Whitney U tests were thus carried out to determine whether the differences were statistically significant (Table 8.34). Results revealed that there were statistically significant differences in the overall LAT between the male group ($M = 7.01$, $SD = 2.92$) and the female group ($M = 11.72$, $SD = 3.33$) in 9th grade ($t(40) = -4.81$, $p = <.001$), and between the male group ($M = 8.42$, $SD = 3.14$) and the female group ($M = 12.51$, $SD = 3.26$) in 11th grade ($t(34) = -3.76$, $p = .001$). However, no such difference was observed between the male group ($M = 9.24$, $SD = 2.58$) and the female group ($M = 10.87$, $SD = 3.19$) in 10th grade ($t(39) = -1.68$, $p = .099$). Regarding the individual prompts, as in the case of the English LAT, results of the independent samples t-tests and Mann-Whitney U tests revealed a number of differences between genders in terms of the individual prompts at each grade. In 9th grade, statistically significant differences between genders were found in all five prompts. This suggests a female advantage across the five lexical domains. In 10th grade, statistically significant differences were found in just two of the five prompts, namely, *Food & Drink* and *Environment & Climate*. Finally, in 11th grade, statistically significant differences were found in all prompts with the exception of one: *Sport & Physical Activities*.

Table 8.33*Descriptive Statistics for Lexical Availability by Gender in French*

Male (<i>n</i> = 16)				9 th Grade	Female (<i>n</i> = 23)			
Min	Max	<i>M</i>	<i>SD</i>	Prompt	Min	Max	<i>M</i>	<i>SD</i>
2	18	7.47	3.53	Animals	4	29	12.70	5.98
1	15	7.05	3.58	Food & Drink	5	25	12.23	4.73
0	15	6.95	3.70	Sport & PA	6	18	11.68	2.93
1	19	8.63	5.37	Env & Climate	2	23	11.83	4.68
0	14	4.95	3.89	Eco & Money	3	18	10.26	4.57
10	70	35.05	14.63	Mean LAT	32	109	57.65	17.46
Male (<i>n</i> = 15)				10 th Grade	Female (<i>n</i> = 26)			
Min	Max	<i>M</i>	<i>SD</i>	Prompt	Min	Max	<i>M</i>	<i>SD</i>
2	12	8.40	2.77	Animals	2	19	8.62	3.87
2	16	8.07	4.31	Food & Drink	3	23	12.38	5.21
9	16	11.47	2.03	Sport & PA	6	17	11.08	2.85
3	16	7.60	3.66	Env & Climate	5	23	12.04	4.60
3	16	10.67	3.71	Eco & Money	2	17	10.27	3.82
25	70	46.20	12.90	Mean LAT	27	83	54.38	15.97
Male (<i>n</i> = 15)				11 th Grade	Female (<i>n</i> = 21)			
Min	Max	<i>M</i>	<i>SD</i>	Prompt	Min	Max	<i>M</i>	<i>SD</i>
4	16	9.27	3.93	Animals	6	18	12.33	3.30
1	16	7.73	4.20	Animals	5	23	13.81	5.19
5	14	9.87	3.33	Food & Drink	4	15	10.38	3.02
2	16	7.93	4.57	Sport & PA	3	21	15.05	5.11
0	12	7.33	3.37	Env & Climate	2	20	11.00	4.69
16	68	42.13	15.70	Eco & Money	23	86	62.57	16.30

Table 8.34*Differences in Lexical Availability in French by Gender*

9 th Grade							
	<i>M</i>		<i>SD</i>		<i>t / z</i>	<i>df</i>	<i>p</i>
	Male	Female	Male	Female			
Animals*	7.47	12.70	3.53	5.98	-3.35		.001
Food & Drink	7.05	12.23	3.58	4.73	-3.89	39	<.001
Sport & Physical Activities*	6.95	11.68	3.70	2.93	-3.75		<.001
Environment & Climate	8.63	11.83	5.37	4.68	-2.05	40	.046
Economy & Money	4.95	10.26	3.89	4.57	-4.00	40	<.001
Mean LAT French	7.01	11.72	2.92	3.33	-4.81	40	<.001
10 th Grade							
	<i>M</i>		<i>SD</i>		<i>T</i>	<i>df</i>	<i>p</i>
	Male	Female	Male	Female			
Animals	8.40	8.62	2.77	3.87	-.18	39	.851
Food & Drink	8.07	12.38	4.31	5.21	-2.7	39	.010
Sport & Physical Activities	11.47	11.08	2.03	2.85	.46	39	.645
Environment & Climate	7.60	12.04	3.66	4.60	-3.1	39	.003
Economy & Money	10.67	10.27	3.71	3.82	.32	39	.748
Mean LAT French	9.24	10.87	2.58	3.19	-1.68	39	.099
11 th Grade							
	<i>M</i>		<i>SD</i>		<i>T</i>	<i>df</i>	<i>p</i>
	Male	Female	Male	Female			
Animals	9.27	12.33	3.93	3.30	-2.53	34	.016
Food & Drink	7.73	13.81	4.20	5.19	-3.73	34	.001
Sport & Physical Activities	9.87	10.38	3.33	3.02	-.48	34	.633
Environment & Climate	7.93	15.05	4.57	5.11	-4.29	34	<.001
Economy & Money	7.33	11.00	3.37	4.69	-2.58	34	.014
Mean LAT French	8.42	12.51	3.14	3.26	-3.76	34	.001

Note. * = non-parametric test used, given that the data were not normally distributed.

These results show that female participants produced a higher number of tokens than the male participants in the French LAT overall, a finding which was statistically significant in 9th and 11th grade but not in 10th grade. Regarding the individual prompts, *Food & Drink* and *Environment & Climate* are the only prompts in which there were statistically significant differences at all levels. Given that the same was found for *Environment & Climate* in the English LAT, this further supports the findings that there is a female advantage in this lexical domain. The prompt *Food & Drink* additionally appears to be an advantageous lexical domain for female students in French, though the same was not found in English. Two prompts revealed statistically significant differences at two levels, namely *Animals* and *Economy & Money* (9th and 11th grade). Finally, statistically significant differences were found for the prompt *Sport & Physical Activities* only in 9th grade.

In order to determine if these findings were consistent with the results of the language level C-test, the normality of the results in French were first assessed for each subgroup of gender in all grades (Table 8.35).

Table 8.35

Normality Tests for the French C-test for Gender

	9 th Grade			10 th Grade			11 th Grade		
	<i>W</i>	<i>df</i>	<i>p</i>	<i>W</i>	<i>df</i>	<i>p</i>	<i>W</i>	<i>df</i>	<i>p</i>
Male	.962	19	.622	.951	15	.533	.828	15	.009
Female	.949	23	.273	.940	26	.138	.958	21	.483

As shown, data were normally distributed at all grades with the exception of 11th grade male students. When these data were involved in the analysis, non-parametric tests were employed. In all other cases, parametric tests were used. Unlike in the case of English, the above findings do not appear to be easily explained by taking into account the participants' language proficiency. In fact, results of independent samples t-tests comparing the results of the French C-test showed that there were statistically significant differences between the male group ($M = 25.95$, $SD = 10.77$) and the female group ($M = 34.13$, $SD = 10.19$) in 9th grade ($t(40) = -2.52$, $p = .016$), and between the male group ($M = 28.40$, $SD = 15.03$) and the female group ($M = 39.50$, $SD = 8.90$) in 10th grade ($t(39) = -2.98$, $p = .005$). Similarly, a Mann-Whitney U test revealed that there was a statistically significant difference between the male group ($M = 37.86$, $SD = 10.56$) and the female group ($M = 43.19$, $SD = 7.71$) in 11th grade ($z = -2.41$, $p = .015$). In all cases, female

participants performed better on the French C-test than their male peers. Thus, female students in 9th and 11th grade had a significantly higher result than their male peers on both the French C-test and the French LAT, and so their performance on the LAT may be attributed to their higher language level. Similar results were found across all levels in English. However, in 10th grade, although female students had a significantly higher result on the French C-test, there was no statistically significant difference between the genders on the French LAT. Thus, at this level, despite having a higher language level, the female students did not outperform their male peers on the LAT to a statistically significant degree.

To summarise and draw together the results of the English and French LATs, Table 8.36 below provides an overview of the gender differences found in each grade. As shown, female students outperformed male students in the overall LAT in both languages across all three grades with just one exception: the French LAT in 10th grade. They also outperformed male students in both languages at all grades for the prompt *Environment & Climate*, suggesting a gender-based advantage in this lexical domain, and across all prompts in French in 9th grade. One additional interesting observation was the fact that, in 11th grade, female students outperformed male students in all prompts of the English LAT with just one exception: *Economy & Money*. This finding is particularly significant given the fact that while male students continued to study economics through English in 11th grade, female students did not. It thus appears that by continuing to receive exposure to content-related vocabulary in the TL, 11th grade male students can bridge the gap between them and their female peers. These results are presented visually in Figure 8.12.

Table 8.36

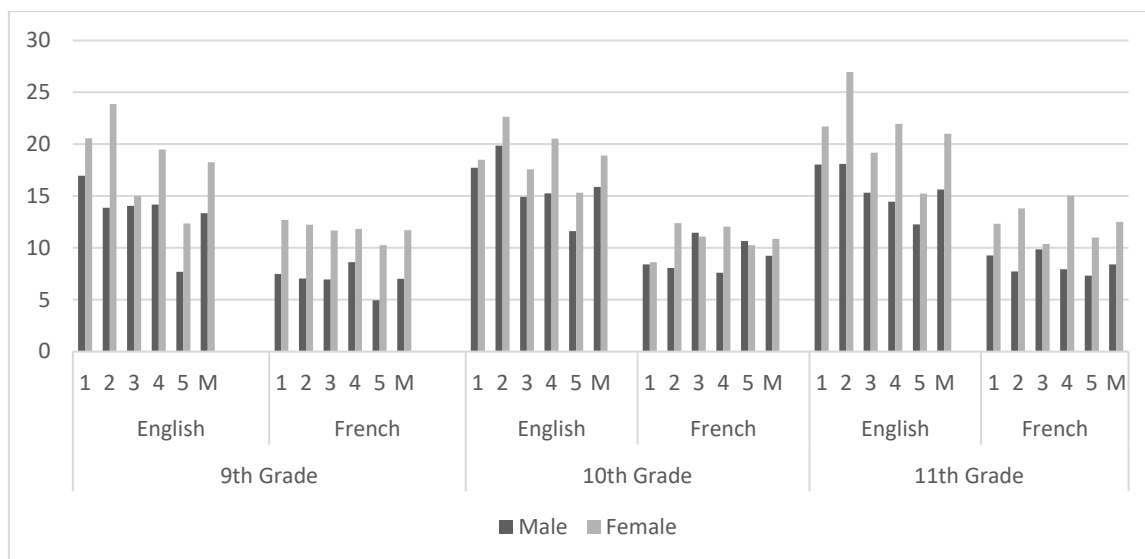
Summary of Statistically Significant Differences between Male and Female Participants within Languages

Prompt		1	2	3	4	5	Mean
9 th Grade	En		✓		✓	✓	✓
	Fr	✓	✓	✓	✓	✓	✓
10 th Grade	En			✓	✓	✓	✓
	Fr		✓		✓		
11 th Grade	En	✓	✓	✓	✓		✓
	Fr	✓	✓		✓	✓	✓

Note. ✓ = statistically significant differences between male and female participants within languages. In all cases, female students produced a higher number of words than male students.

Figure 8.12

Differences in LAT across Gender



Note. 1 = *Animals*, 2 = *Food & Drink*, 3 = *Sport and Physical Activities*, 4 = *Environment & Climate*, 5 = *Economy & Money*, M = Mean LAT.

In order to determine whether there were cross-sectional (9th and 10th grade) differences between male and female participants in the LATs at each testing period, independent samples t-tests / Mann-Whitney U tests, were carried out, given that different participants were being compared. The results for each group were then compared in order to determine differences between the genders. This was again done first for English and then for French. Repeated measures ANOVA were not used in this analysis, as was done in the rest of the cross-sectional and longitudinal analyses of gender, given that different groups of participants were being compared; as a result, two between-subjects factors were being investigated (grade and gender) while the within-subjects factor contained just one level.

Firstly, regarding cross-sectional differences in English (Table 8.37), independent samples t-tests and Mann-Whitney U tests comparing male and female participants in 9th and 10th grade revealed statistically significant differences in the prompt *Food & Drink* for male students and only for the prompt *Economy & Money* for female students. While 10th grade male students produced a statistically significant higher number of words in the general domain of *Food & Drink* than 9th grade male students, 10th grade female students produced a statistically significant higher number of words in the content specific domain of *Economy & Money* than 9th grade female students. In the case of the latter, this may be explained by the fact that a number of students in the female group would have

begun to study economics in 10th grade, and thus expanded their vocabulary in this area to a greater extent. However, it is unclear why no such difference was found in this prompt for male students, given that similar percentages of male and female students had enrolled in economics in 10th grade (33% male students and 38% female students). In order to determine whether these differences were attributable to language level, independent samples t-tests were carried out to compare participants' performance on the English C-test in each grade. Results revealed a statistically significant difference in the English C-test for male students ($t(31) = -3.74, p = .001$), with an improvement from 9th grade ($M = 22.28, SD = 8.40$) to 10th grade ($M = 33.80, SD = 9.27$). However, no statistically significant difference was found in the English C-test for female students ($t(47) = .089, p = .929$), with a very slight decrease from 9th grade ($M = 34.52, SD = 12.47$) to 10th grade ($M = 34.19, SD = 13.31$).

Secondly, regarding cross-sectional differences in French (Table 8.38), independent samples t-tests and Mann-Whitney U tests comparing male and female participants in 9th and 10th grade showed that there were statistically significant differences in the prompts *Sport & Physical Activities* and *Economy & Money* and in the overall mean for male students, and in the prompt *Animals* for female students. This indicates that 10th grade male students produced a statistically significant higher number of words than 9th grade students both overall and in the prompts *Sport & Physical Activities* and *Economy & Money*, which incidentally were related not to their French CLIL classes but rather to their English CLIL classes. A more curious finding is that of the female students, where in fact the younger, 9th grade students produced a statistically significant higher number of words than 10th grade student for the general prompt *Animals*. While not statistically significant, these younger female students also produced a higher number of words overall and in the prompt *Sport & Physical Activities*. To determine whether these differences were attributable to language level, independent samples t-tests were carried out to compare participants' performance on the French C-test in each grade. As in the case of the English C-test, differences varied depending on gender. A statistically significant difference was found in the English C-test for male students ($t(32) = -2.28, p = .029$), with an improvement from 9th grade ($M = 25.95, SD = 10.77$) to 10th grade ($M = 34.93, SD = 12.14$). However, no statistically significant difference was found in the English C-test for female students ($t(47) = -.474, p = .638$), with a slight increase from 9th grade ($M = 34.13, SD = 10.19$) to 10th grade ($M = 35.73, SD = 13.03$).

Table 8.37*Cross-Sectional Differences in the English LAT by Gender*

9 th → 10 th grade: Male							
Prompt	<i>M</i>		<i>SD</i>		<i>t / z</i>	<i>df</i>	<i>p</i>
	9 th	10 th	9 th	10 th			
Animals*	16.94	17.73	8.31	4.49	-.317		.770
Food & Drink	13.88	19.87	8.17	4.73	-2.47	29	.019
Sport & Physical Activities	14.06	14.93	7.10	2.76	-.455	19.68	.654
Environment & Climate	14.19	15.27	8.47	6.19	-.403	29	.690
Economy & Money	7.69	11.60	6.61	4.71	-1.88	29	.069
Mean LAT English	13.35	15.88	6.90	3.41	-1.27	29	.211
9 th → 10 th grade: Female							
Prompt	<i>M</i>		<i>SD</i>		<i>t</i>	<i>df</i>	<i>p</i>
	9 th	10 th	9 th	10 th			
Animals	20.57	18.50	5.085	5.86	1.32	47	.193
Food & Drink	23.87	22.62	5.426	6.31	.748	47	.458
Sport & Physical Activities	15.00	17.58	5.139	5.04	-1.76	47	.084
Environment & Climate	19.48	20.54	4.794	6.29	-.668	47	.508
Economy & Money	12.35	15.31	4.361	4.83	-2.25	47	.029
Mean LAT English	18.25	18.90	3.96	4.96	-.514	47	.610

Note. * = non-parametric test used, given that the data were not normally distributed.

Table 8.38*Cross-Sectional Differences in the French LAT by Gender*

9 th → 10 th grade: Male							
Prompt	<i>M</i>		<i>SD</i>		<i>t / z</i>	<i>df</i>	<i>p</i>
	9 th	10 th	9 th	10 th			
Animals	7.47	8.40	3.53	2.77	-.832	32	.411
Food & Drink	7.05	8.07	3.58	4.31	-.749	32	.459
Sport & Physical Activities*	6.95	11.47	3.70	2.03	-3.44		<.001
Environment & Climate	8.63	7.60	5.377	3.66	.635	32	.530
Economy & Money	4.95	10.67	3.89	3.71	-4.33	32	<.001
Mean LAT French	7.01	9.24	2.92	2.58	-2.32	32	.027
9 th → 10 th grade: Female							
Prompt	<i>M</i>		<i>SD</i>		<i>t / z</i>	<i>df</i>	<i>p</i>
	9 th	10 th	9 th	10 th			
Animals*	12.70	8.62	5.98	3.87	-2.42		.015
Food & Drink	12.23	12.38	4.73	5.21	-.109	46	.914
Sport & Physical Activities	11.68	11.08	2.93	2.85	.722	46	.474
Environment & Climate	11.83	12.04	4.68	4.60	-.160	47	.874
Economy & Money	10.26	10.27	4.57	3.82	-.007	47	.994
Mean LAT French	11.72	10.87	3.33	3.19	.911	47	.367

Note. * = non-parametric test used, given that the data were not normally distributed.

In order to determine whether there were longitudinal differences across gender in the English and French LATs, two-way mixed ANOVAs were carried out comparing male and female participants in 10th and 11th grade, first in English and then in French.

Firstly, regarding longitudinal differences in English, analyses were carried out on male and female participants' performance on the LAT (Table 8.39). Results indicated that there was a statistically significant main effect of gender on the overall LAT ($F(1, 30) = 11.61, p = .002, \eta_p^2 = .279$), with female participants ($M = 20.46$) performing better than male participants ($M = 15.75$) overall. There was, however, no statistically significant main effect of time ($F(1, 30) = 1.40, p = .244, \eta_p^2 = .045$), though means were higher in 11th grade ($M = 18.37$) than in 10th grade ($M = 17.84$). There was no statistically significant interaction between gender and time in English ($F(1, 30) = 2.68, p = .112, \eta_p^2 = .082$); male participants' score decreased slightly from 10th grade ($M = 15.85$) to 11th

grade ($M = 15.65$), while female participants' score increased from 10th grade ($M = 19.83$) to 11th grade ($M = 21.08$). Regarding the individual prompts, results indicated that there was a statistically significant main effect of gender on four out of the five prompts, with the exception of the prompt *Animals*. In all cases, female participants produced a higher number of tokens across the two data collections. A statistically significant main effect of time was found only in the case of the prompt *Animals*, for which participants produced a higher number of tokens in 11th grade than in 10th grade. There was no statistically significant interaction between gender and time for the majority of the prompts, with just one exception: *Food & Drink* ($F(1, 30) = 13.98, p = .001, \eta_p^2 = .318$). Result showed that male participants' score on this prompt decreased quite a bit from 10th grade ($M = 20.35$) to 11th grade ($M = 17.92$) while females score increased a great deal from 10th grade ($M = 23.55$) to 11th grade ($M = 27.05$). There was consequently a much larger differences between male and female participants in 11th grade than in 10th grade in this prompt. Regarding language level, results indicated that there was a significant main effect of gender on the English C-test at each time ($F(1, 30) = 7.94, p = .008, \eta_p^2 = .209$), with female participants ($M = 40.22$) performing better than male participants ($M = 29.07$) overall. There was also a significant main effect of time ($F(1, 30) = 7.09, p = .012, \eta_p^2 = .191$), with participants improving from 10th grade ($M = 33.04$) to 11th grade ($M = 36.25$). However, there was no statistically significant interaction between time and gender ($F(1, 30) = .292, p = .593, \eta_p^2 = .010$); male participants' language level increased from 10th grade ($M = 27.14$) to 11th grade ($M = 31.00$), and female participants' language level increased from 10th grade ($M = 38.94$) to 11th grade ($M = 41.50$).

Secondly, regarding longitudinal differences in French, two-way mixed ANOVAs were carried out on male and female participants' performance on the LAT (Table 8.40). Results indicated that there was a statistically significant main effect of gender on the overall LAT ($F(1, 28) = 11.95, p = .002, \eta_p^2 = .299$), with female participants ($M = 12.44$) again performing better than male participants ($M = 8.81$) overall. There was, however, no statistically significant main effect of time ($F(1, 28) = 2.83, p = .103, \eta_p^2 = .092$), where means were just slightly higher in 11th grade ($M = 10.90$) than in 10th grade ($M = 10.36$). There was a statistically significant interaction between gender and time in French ($F(1, 28) = 7.99, p = .009, \eta_p^2 = .222$); male participants' score decreased slightly from 10th grade ($M = 9.00$) to 11th grade ($M = 8.63$), while female participants' score increased from 10th grade ($M = 11.72$) to 11th grade ($M = 13.16$). A much larger difference was thus found between male and female participants in 11th grade than in 10th grade in overall

French LAT. Regarding the individual prompts, results indicated that there was a significant main effect of gender for just one of the five prompts: *Environment & Climate* ($F(1, 28) = 18.74, p = < .001, \eta_p^2 = .401$), which is notably the prompt which was included to tap into the participants content-related vocabulary from their French CLIL class. In this prompt, and all others, female participants produced a higher number of tokens across the two data collections. A statistically significant main effect of time was found for *Animals, Environment & Climate* and *Economy & Money*. While for the former two, participants produced a higher number of tokens in 11th grade than in 10th grade, they produced fewer tokens in 11th grade than 10th grade for the prompt *Economy & Money*. There was no statistically significant interaction between gender and time for the majority of the prompts, with just one exception: *Economy & Money* ($F(1, 28) = 6.26, p = .018, \eta_p^2 = .183$). Results showed that male participants' score on this prompt decreased quite a bit from 10th grade ($M = 10.58$) to 11th grade ($M = 7.08$) while females score increased very slightly from 10th grade ($M = 11.16$) to 11th grade ($M = 11.50$). There was thus again a much larger differences between male and female participants in 11th grade than in 10th grade. Regarding language level, results indicated that there was no significant main effect of gender on the French C-test at each time ($F(1, 28) = 3.47, p = .073, \eta_p^2 = .110$), though female participants ($M = 41.19$) performed better than male participants ($M = 36.41$) overall. There was, however, a statistically significant main effect of time ($F(1, 28) = 4.44, p = .049, \eta_p^2 = .131$), with participants improving from 10th grade ($M = 35.58$) to 11th grade ($M = 42.02$). There was no statistically significant interaction between gender and time ($F(1, 28) = .091, p = .765, \eta_p^2 = .003$); male participants' French language level increased from 10th grade ($M = 33.66$) to 11th grade ($M = 39.16$), and female participants' French language level increased from 10th grade ($M = 37.50$) to 11th grade ($M = 44.88$).

Table 8.39*Longitudinal Differences in the English LAT by Gender*

Gender							
	<i>M</i>				<i>F</i>	<i>p</i>	η_p^2
	Male		Female				
1	18.14		20.44		1.75	.195	.055
2	19.14		25.30		11.36	.002	.275
3	15.03		18.77		7.30	.011	.196
4	14.85		22.13		13.12	.001	.304
5	11.60		15.63		6.99	.013	.189
Mean	15.75		20.46		11.61	.002	.279
Time							
	<i>M</i>				<i>F</i>	<i>p</i>	η_p^2
	10 th Grade		11 th Grade				
1	18.54		20.04		4.24	.048	.124
2	21.95		22.49		.457	.504	.015
3	16.93		16.87		.009	.927	.000
4	18.18		18.81		.556	.462	.018
5	13.59		13.64		.005	.946	.000
Mean	17.84		18.37		1.40	.244	.045
Gender*Time							
	<i>M</i>				<i>F</i>	<i>p</i>	η_p^2
	10 th Grade		11 th Grade				
	Male	Female	Male	Female			
1	17.92	19.16	18.35	21.72	2.15	.152	.067
2	20.35	23.55	17.92	27.05	13.98	.001	.318
3	14.92	18.94	15.14	18.61	.181	.674	.006
4	14.92	21.44	14.78	22.83	.840	.367	.027
5	11.14	16.05	12.07	15.22	1.57	.219	.050
Mean	15.85	19.83	15.65	21.08	2.68	.112	.082

Note. 1 = Animals, 2 = Food & Drink, 3 = Sport and Physical Activities, 4 = Environment & Climate, 5 = Economy & Money, M = Mean LAT.

Table 8.40*Longitudinal Differences in the French LAT by Gender*

Gender								
	<i>M</i>			<i>F</i>	<i>p</i>	ηp^2		
	Male		Female					
1	8.87		10.88	3.07	.091	.099		
2	7.58		14.16	1.16	.289	.040		
3	10.45		11.38	1.18	.285	.041		
4	8.33		14.44	18.74	<.001	.401		
5	8.83		11.33	3.68	.065	.116		
Mean	8.81		12.44	11.95	.002	.299		
Time								
	<i>M</i>			<i>F</i>	<i>p</i>	ηp^2		
	10 th Grade		11 th Grade					
1	8.65		11.11	14.29	.001	.338		
2	10.51		11.23	1.16	.289	.040		
3	11.38		10.45	3.29	.080	.105		
4	10.37		12.40	8.94	.006	.242		
5	10.87		9.29	4.27	.048	.132		
Mean	10.36		10.90	2.83	.103	.092		
Gender*Time								
	<i>M</i>				<i>F</i>	<i>p</i>	ηp^2	
	10 th Grade		11 th Grade					
	Male	Female	Male	Female				
1	8.08	9.22	9.66	12.55	1.81	.189	.061	
2	7.58	13.44	7.58	14.88	1.16	.289	.040	
3	10.83	11.94	10.08	10.83	.124	.727	.004	
4	7.91	12.83	8.75	16.05	3.10	.089	.100	
5	10.58	11.16	7.08	11.50	6.26	.018	.183	
Mean	9.00	11.72	8.63	13.16	7.99	.009	.222	

Note. 1 = Animals, 2 = Food & Drink, 3 = Sport and Physical Activities, 4 = Environment & Climate, 5 = Economy & Money, M = Mean LAT.

In summary, regarding cross-sectional differences, results indicated practically no similarity between male and female students. As summarised in Table 8.41, in no cases were there statistically significant difference from one grade to the next in both male and female students. All differences were observed only for male students or female students, indicating that they have not progressed in a similar way. From 9th to 10th grade, 10th grade male students performed better in the prompt *Food & Drink* in English and in the prompts *Sport & Physical Activities* and *Economy & Money* and overall in French, while 9th grade female students performed better in the prompt *Animals* in French and 10th grade female students performed better in the prompt *Economy & Money* in English.

Table 8.41

Summary of Cross-Sectional Differences in LAT and C-test by Gender within Languages

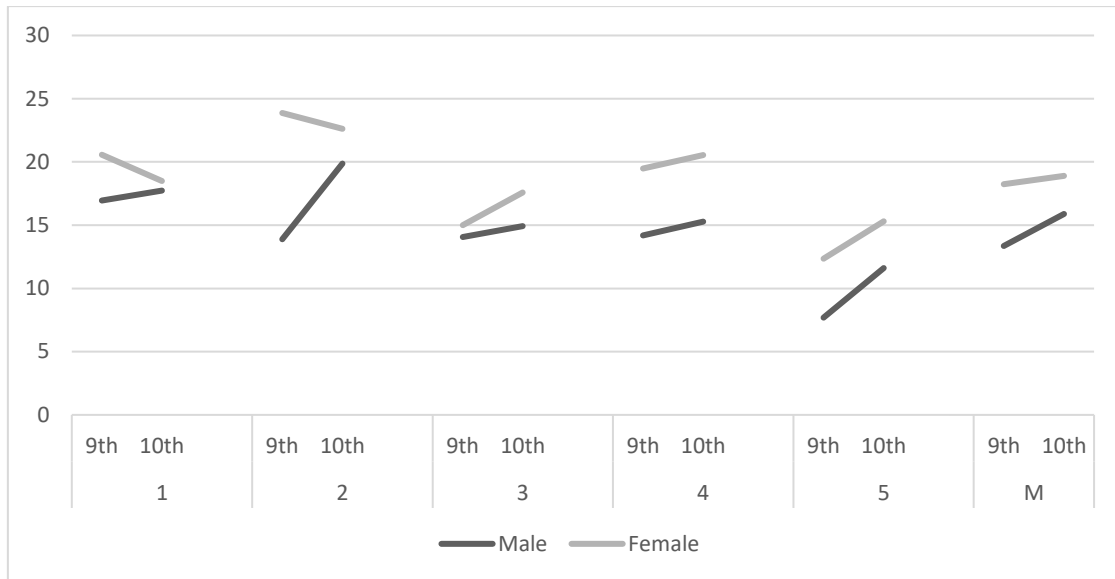
Prompt		1	2	3	4	5	Mean	C-test
9 th → 10 th English	Male		✓					✓
	Female					✓		
9 th → 10 th French	Male			✓		✓	✓	✓
	Female	✓						

Note. ✓ = statistically significant differences between participants in from one grade to the next. In general, the more advanced grade produced a higher number of tokens, with the exception of Prompt 1 in French, where female students produced more tokens in 9th grade than those in 10th grade.

These results suggest that while improvements are seen from one grade to the next, there is a lack of similarity in male and female students, with each group progressing to a greater or lesser extent in different prompts and languages. In addition, analysis of language level indicated that while male students improved from one grade to the next, the same was not true for female students. This finding could potentially explain the greater improvement shown by male students in the French LAT. Figures 8.13-8.14 provide a visual representation of these results.

Figure 8.13

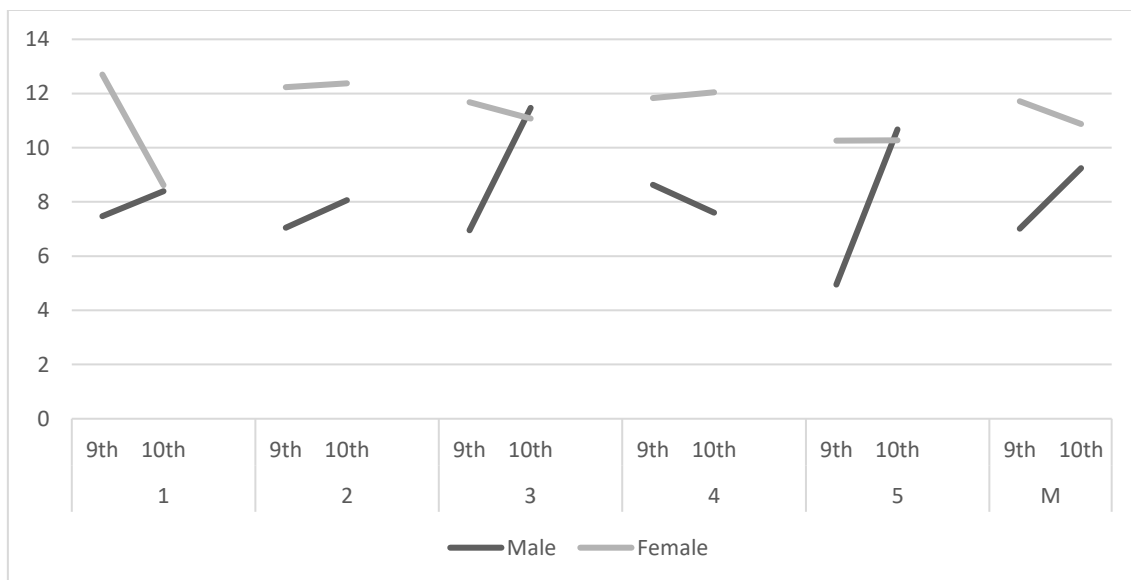
Cross-Sectional Differences in the English LAT by Gender



Note. 1 = Animals, 2 = Food & Drink, 3 = Sport and Physical Activities, 4 = Environment & Climate, 5 = Economy & Money, M = Mean LAT, C = C-test.

Figure 8.14

Cross-Sectional Differences in the French LAT by Gender



Note. 1 = Animals, 2 = Food & Drink, 3 = Sport and Physical Activities, 4 = Environment & Climate, 5 = Economy & Money, M = Mean LAT.

Regarding longitudinal differences, the two-way mixed ANOVAs revealed interesting differences between male and female participants at each data collection in English and French, as summarised in Table 8.42.

Table 8.42

Summary of Longitudinal Differences in LAT and C-test by Gender within Languages

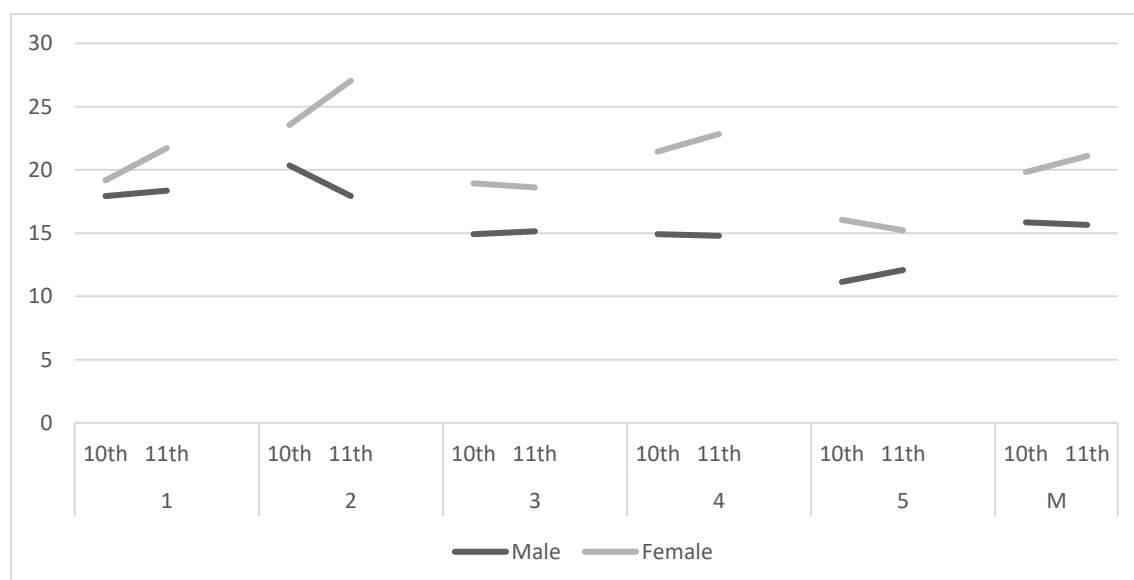
Gender		
	English	French
Animals		
Food & Drink	✓	
Sport & Physical Activity	✓	
Environment & Climate	✓	✓
Economy & Money	✓	
Mean LAT	✓	✓
C-test	✓	
Time		
	English	French
Animals	✓	✓
Food & Drink		
Sport & Physical Activity		
Environment & Climate		✓
Economy & Money		✓
Mean LAT		
C-test	✓	✓
Gender*Time		
	English	French
Animals		
Food & Drink	✓	
Sport & Physical Activity		
Environment & Climate		
Economy & Money		✓
Mean LAT		✓
C-test		

Firstly, regarding gender, the longitudinal analysis indicated that while there was significant main effect of gender on the C-tests, the overall LAT and all prompts except *Animals* in English, there was significant main effect of gender only on the overall LAT and the prompt *Environment & Climate* in French. This suggests that there is a larger difference between male and female participants across the two data collections in

English than in French. In all cases, female participants had higher means than the male participants. Secondly, regarding time, the longitudinal analysis revealed statistically significant differences in the C-test and the prompt *Animals* in English, and in the C-test and the prompts *Animals*, *Environment & Climate* and *Economy & Money* in French. This indicates that both male and female participants show a clearer difference from 10th to 11th grade in French as compared with English. This may be due to the fact that, given that the students' English language level was much higher than their French language level, there may be less room for improvement in English than in French. Finally, there was a statistically significant interaction between gender and time only in the prompt *Food & Drink* in English, whereas there was a statistically significant interaction between gender and time in the overall LAT and the prompt *Economy & Money* in French. In both languages, male participants' scores decreased from 10th grade to 11th grade, while females' scores increased from 10th to 11th grade. As a result, a much larger difference was found between male and female participants in 11th grade than in 10th grade. Figures 8.15-8.16 provide a visual representation of these results.

Figure 8.15

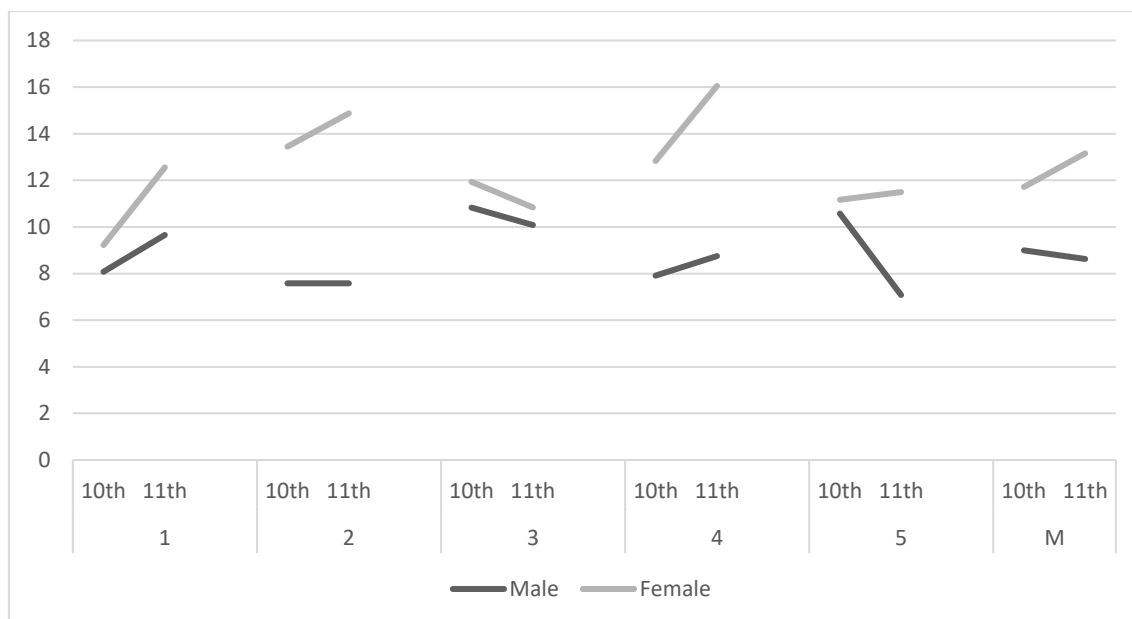
Longitudinal Differences in the English LAT by Gender



Note. 1 = *Animals*, 2 = *Food & Drink*, 3 = *Sport and Physical Activities*, 4 = *Environment & Climate*, 5 = *Economy & Money*, M = Mean LAT.

Figure 8.16

Longitudinal Differences in the French LAT by Gender



Note. 1 = *Animals*, 2 = *Food & Drink*, 3 = *Sport and Physical Activities*, 4 = *Environment & Climate*, 5 = *Economy & Money*, M = Mean LAT.

8.3.2. Qualitative Differences in Gender and LA

Research question 3.2 aimed to determine whether there were qualitative differences between male and female students in terms of the frequency of first word responses for each prompt, the most and least productive prompts, and the lexical sophistication of each language based on the non-shared words of participants as well as the number of infrequent words in the production of each prompt.

Firstly, with regard to the most frequent word for each prompt, Wordsmith Tools was used to analyse the most frequent first word responses retrieved by each group for each prompt, and the number and percentage of participants that retrieved the response, both for English (Table 8.43) and French (Table 8.44). As can be seen, results showed a great deal of similarity between both groups in the prompts *Animals*, *Food & Drink*, *Environment & Climate* in both English and French, with the words “dog”, “water” and “sun” being the most frequent in both languages for both male and female participants in all three grades. However, with regard to *Sport & Environment*, the most frequent first word in both English and French for male participants was “football” whereas for female participants variation was found across grades and languages.

Table 8.43*Most Frequent First Words in English by Gender*

9 th grade				
	Male		Female	
Prompt	Lexical Unit	N of informants	Lexical Unit	N of informants
1	<i>Dog</i>	9 (56.3%)	<i>Dog</i>	7 (30.43%)
2	<i>Water</i>	5 (33.3%)	<i>Water</i>	10 (43.5%)
3	<i>Football</i>	8 (50%)	<i>Tennis</i>	7 (30.4%)
4	<i>Sun</i>	6 (37.5%)	<i>Sun</i>	6 (26.1%)
5	<i>Euro</i>	2 (15.4%)	<i>Dollar</i>	6 (26.1%)
	<i>Footballplayer</i>	2 (15.4%)		
	<i>Money</i>	2 (15.4%)		
10 th grade				
	Male		Female	
Prompt	Lexical Unit	N of informants	Lexical Unit	N of informants
1	<i>Dog</i>	9 (60%)	<i>Dog</i>	14 (53.6%)
2	<i>Ketchup</i>	2 (13.3%)	<i>Water</i>	10 (38.5%)
	<i>Water</i>	2 (13.3%)		
3	<i>Football</i>	10 (66.7%)	<i>Basketball</i>	5 (19.2%)
4	<i>Sun</i>	5 (33.3%)	<i>Sun</i>	5 (19.2%)
5	<i>Dollar</i>	3 (20%)	<i>Economy</i>	8 (30.8%)
11 th grade				
	Male		Female	
Prompt	Lexical Unit	N of informants	Lexical Unit	N of informants
1	<i>Dog</i>	7 (36.8%)	<i>Dog</i>	6 (28.6%)
2	<i>Water</i>	5 (26.3%)	<i>Water</i>	7 (33.3%)
3	<i>Football</i>	11 (57.9%)	<i>Basketball</i>	5 (23.8%)
			<i>Football</i>	5 (23.8%)
4	<i>Sun</i>	4 (21.1%)	<i>Sun</i>	5 (23.8%)
5	<i>Money</i>	4 (21.1%)	<i>Money</i>	7 (33.3%)

Note. 1 = Animals, 2 = Food & Drink, 3 = Sport and Physical Activities, 4 = Environment & Climate, 5 = Economy & Money, M = Mean LAT.

Table 8.44*Most Frequent First Words in French by Gender*

9 th grade				
	Male		Female	
Prompt	Lexical Unit	N of informants	Lexical Unit	N of informants
1	<i>Chien</i>	11 (57.9%)	<i>Chien</i>	5 (21.7%)
2	<i>Eau</i>	5 (26.3%)	<i>Eau</i>	6 (27.3%)
3	<i>Football</i>	13 (72.2%)	<i>Natation</i>	6 (27.3%)
4	<i>Soleil</i>	14 (73.7%)	<i>Soleil</i>	9 (39.1%)
5	<i>Argent</i>	3 (16.7%)	<i>Euro</i>	7 (30.4%)
	<i>Euro</i>	3 (16.7%)		
10 th grade				
	Male		Female	
Prompt	Lexical Unit	N of informants	Lexical Unit	N of informants
1	<i>Chien</i>	7 (46.7%)	<i>Chien</i>	13 (50%)
2	<i>Eau</i>	7 (46.7%)	<i>Eau</i>	10 (38.5%)
3	<i>Football/Foot</i>	9 (60%)	<i>Basket</i>	9 (34.6%)
4	<i>Soleil</i>	6 (40%)	<i>Soleil</i>	7 (26.9%)
5	<i>Économie</i>	5 (33.3%)	<i>Argent</i>	7 (26.9%)
11 th grade				
	Male		Female	
Prompt	Lexical Unit	N of informants	Lexical Unit	N of informants
1	<i>Chien</i>	4 (26.7%)	<i>Chien</i>	9 (42.9%)
2	<i>Eau</i>	9 (60%)	<i>Eau</i>	9 (42.9%)
3	<i>Football/Foot</i>	7 (46.7%)	<i>Basket</i>	5 (28.8%)
4	<i>Soleil</i>	7 (46.7%)	<i>Soleil</i>	7 (33.3%)
5	<i>Argent</i>	8 (57.1%)	<i>Argent</i>	10 (47.6%)

Note. 1 = Animals, 2 = Food & Drink, 3 = Sport and Physical Activities, 4 = Environment & Climate, 5 = Economy & Money, M = Mean LAT.

For 9th grade females it was “tennis” in English and “swimming” in French; in 10th grade it was “basketball” in both English and French, and in 11th grade it was “basketball” / “football” in English and “basketball” in French. These differences may be attributed to the interests of the participants, given that football is often the chosen sport by male students while basketball is more popular among female students. With regard to *Economy & Money*, differences arose again both across grades and languages. In 9th grade, more variability was noted among male students, where “euro” / “footballplayer” / “money” were the most common first words in English and “dollar” / “money” were the most common in French. For female students, “dollar” was the most common in English and “euro” was the most common in French. In 10th grade, the most frequent first word in English for male students was “dollar” whereas for female students it was “economy”. In French the most frequent word for males was “economy” whereas for females it was “money”. Finally, in 11th grade, more similarity was observed, as “money” was the most common for both male and female students in English and in French. Despite the variety of responses for this prompt, those that were most common tended to appear in the prompt name itself (“economy” / “money”) or to be a type of currency (“euro” / “dollar”).

Secondly, the prompts were analysed in order to determine the ranking of the most and least productive prompts in English (Table 8.45) and in French (Table 8.46). Regarding English, results showed a great deal of similarity between male and female participants, with *Food & Drink* generally occupying the first position, with the exception of 9th grade male students, and *Economy & Money* occupying the last in all grades. *Sport & Physical Activities* was similarly often one of the least productive, generally ranking in fourth position. However, regarding French, while similarities were seen in the 9th grade male and female students, there was a great deal of variety in the ranking in 10th and 11th grade. Some trends included the fact that *Sport & Physical Activities* ranked first for male students in both grades, whereas for females it ranked third in 10th grade and last in 11th grade. On the other hand, *Environment & Climate* ranked higher for female students, ranking second in 10th grade and first in 11th grade, than male students, where it ranked last in 10th grade and third in 11th grade. This could again perhaps be attributed to student interest in the topic. In order to investigate these differences further, an analysis was carried out on the presence of cognates, as in the analysis of research question 1.2. VocabProfile was used to determine the percentage of French cognates in the English LAT (Table 8.47) and the percentage of English cognates in the French LAT (Table 8.48) for each group.

Table 8.45*Ranking of Most and Least Productive Prompts in English by Gender*

9 th grade				
Rank	Male	Tokens	Female	Tokens
1	Animals	271	Food & Drink	549
2	Environment & Climate	227	Animals	473
3	Sport & Physical Activities	225	Environment & Climate	448
4	Food & Drink	222	Sport & Physical Activities	345
5	Economy & Money	123	Economy & Money	284
10 th grade				
Rank	Male	Tokens	Female	Tokens
1	Food & Drink	298	Food & Drink	588
2	Animals	266	Environment & Climate	534
3	Environment & Climate	229	Animals	481
4	Sport & Physical Activities	224	Sport & Physical Activities	457
5	Economy & Money	174	Economy & Money	398
11 th grade				
Rank	Male	Tokens	Female	Tokens
1	Food & Drink	344	Food & Drink	566
2	Animals	343	Environment & Climate	461
3	Sport & Physical Activities	291	Animals	456
4	Environment & Climate	275	Sport & Physical Activities	403
5	Economy & Money	233	Economy & Money	320

Table 8.46*Ranking of Most and Least Productive Prompts in French by Gender*

9 th grade				
Rank	Male	Tokens	Female	Tokens
1	Environment & Climate	164	Animals	292
2	Animals	142	Environment & Climate	272
3	Food & Drink	134	Food & Drink	269
4	Sport & Physical Activities	132	Sport & Physical Activities	257
5	Economy & Money	94	Economy & Money	236
10 th grade				
Rank	Male	Tokens	Female	Tokens
1	Sport & Physical Activities	172	Food & Drink	322
2	Economy & Money	160	Environment & Climate	313
3	Animals	126	Sport & Physical Activities	288
4	Food & Drink	121	Economy & Money	267
5	Environment & Climate	114	Animals	224
11 th grade				
Rank	Male	Tokens	Female	Tokens
1	Sport & Physical Activities	148	Environment & Climate	316
2	Animals	139	Food & Drink	290
3	Environment & Climate	119	Animals	259
4	Food & Drink	116	Economy & Money	231
5	Economy & Money	110	Sport & Physical Activities	218

Table 8.47*Cognates in the English LAT by Gender*

	9 th grade		10 th grade		11 th grade	
	Prompt	Cognates	Prompt	Cognates	Prompt	Cognates
Male	5	54%	5	57%	5	57%
	2	45%	2	49%	2	48%
	3	37%	3	41%	3	44%
	1	26%	1	27%	1	29%
	4	24%	4	22%	4	20%
Female	2	46%	5	46%	2	43%
	3	39%	2	42%	3	36%
	5	38%	3	38%	5	36%
	1	25%	1	23%	1	25%
	4	18%	4	20%	4	21%

Note. Content-relevant prompts for each language are marked in bold. Prompt 1 = *Animals*, Prompt 2 = *Food & Drink*, Prompt 3 = *Sport and Physical Activities*, Prompt 4 = *Environment & Climate*, Prompt 5 = *Economy & Money*.

In the English LAT, the first observation was the remarkable similarity in cognateness across prompts in both male and female students. For male students at all grades and females in 10th grade, *Economy & Money* contained the highest percentage of French cognates, followed by *Food & Drink*, *Sport and Physical Activities*, *Animals*, and finally *Environment & Climate*. Female students in 9th and 11th grade also had a very similar pattern, the only difference being that *Economy & Money* had the third highest percentage of French cognates as opposed to the first. There did not seem to be a relationship between the ranking and the percentage of cognates. While *Food & Drink* generally occupied the first position in the ranking, there was generally fewer instance of cognates; *Economy & Money* ranked last in all grades, despite generally having the highest percentage of cognates; and *Sport & Physical Activities*, though it was generally one of the least productive, ranked around third in terms of cognateness. This indicates that male and female students produce very similar results in terms of the most and least productive prompts in the English LAT, though this does not seem to be affected by reliance on cognates.

Table 8.48*Cognates in the French LAT by Gender*

	9 th grade		10 th grade		11 th grade	
	Prompt	Cognates	Prompt	Cognates	Prompt	Cognates
Male	3	87%	5	78%	3	91%
	5	72%	3	74%	5	71%
	2	31%	4	33%	2	31%
	1	27%	2	20%	1	16%
	4	26%	1	6%	4	14%
Female	3	84%	3	76%	3	73%
	5	75%	5	65%	5	63%
	2	40%	2	47%	2	40%
	4	29%	4	31%	4	29%
	1	26%	1	16%	1	14%

Note. Content-relevant prompts for each language are marked in bold. Prompt 1 = *Animals*, Prompt 2 = *Food & Drink*, Prompt 3 = *Sport and Physical Activities*, Prompt 4 = *Environment & Climate*, Prompt 5 = *Economy & Money*.

In the French LAT, as in the case of English, there was a great deal of similarity of cognateness between male and female learners in each grade. In female groups in all grades, *Sport and Physical Activities* contained the highest percentage of English, followed by *Economy & Money*, *Food & Drink*, *Environment & Climate* and *Animals*. Male students in 9th and 11th had almost the same the same results, with the prompt *Environment & Climate* containing just slightly fewer cognates than *Animals*. The largest difference was seen between male and female students in 10th grade, where female students' responses contained a higher percentage of cognates in the general prompts *Animals* (Males: 6%, Females: 16%) and *Food & Drink* (Males: 20%, Females: 47%). Male students on the other hand, had a higher percentage of cognates in the prompt *Economy & Money* (Males: 78%, Females: 65%). As noted above, in the French LAT, unlike in the English LAT, there was much more variety in the ranking of male and female students in 10th and 11th grade, including the fact that *Sport and Physical Activities* ranked higher for male students than female students. However, there does not seem to be a relationship between these differences in ranking and the level of cognateness.

Finally, analyses were carried out to examine lexical sophistication to compare

male and female participants based on the non-shared words of participants as well as the number of infrequent words in the production of each prompt.

With regard to non-shared words, each prompt was analysed in turn to determine the number of words which were unique to one participant, and the percentage of these non-shared words. For example, in the prompt *Economy & Money* in English, male students in 9th grade had a total of 64 non-shared words out of a total of 123 words, meaning that 52% of the words produced by the participants were non-shared words. The prompts were then ranked to examine the prompts with the most and least non-shared words for English (Table 8.49) and French (Table 8.50). Regarding English, there is clear similarity in the ranking in male and female students in each grade, with *Economy & Money* having the highest percentage of non-shared words, *Environment & Climate* and *Sport & Physical Activities* having the second or third highest, and *Food & Drink* and *Animals* having the least. Notably, however, male students in each grade had a higher percentage of non-shared words than their female peers in almost all cases, with the exception of *Animals* in 10th grade and *Sport & Physical Activities* in 10th and 11th grade. As noted in the quantitative analysis, male students produced a statistically significant lower number of words than female students in the category *Animals* in 11th grade, *Food & Drink* in 9th and 11th grade, *Sport & Physical Activities* in 10th and 11th grade and *Economy & Money* in 9th and 10th grade. However, despite these quantitative advantages on the part of female students, the above results suggest that, qualitatively, male students have a higher level of lexical sophistication in terms of non-shared words in the majority of the categories under investigation.

Table 8.49*Non-shared Words in English by Gender*

9 th grade				
Rank	Male	Non-shared words	Female	Non-shared words
1	Economy & Money	64/123 52%	Economy & Money	73/284 18.3%
2	Sport & Physical Activities	68/225 30.2%	Sport & Physical Activities	61/345 13.3%
3	Environment & Climate	63/227 27.8%	Environment & Climate	60/448 11.2%
4	Food & Drink	46/222 20.7%	Food & Drink	48/549 8.2%
5	Animals	47/271 17.3%	Animals	33/473 6.9%
10 th grade				
Rank	Male	Non-shared words	Female	Non-shared words
1	Economy & Money	60/174 34.5%	Economy & Money	114/398 28.6%
2	Environment & Climate	57/229 24.9%	Sport & Physical Activities	95/457 20.8%
3	Sport & Physical Activities	41/224 18.3%	Environment & Climate	93/534 17.4%
4	Food & Drink	39/298 13.1%	Animals	77/481 16%
5	Animals	32/266 12%	Food & Drink	76/588 12.9%
11 th grade				
Rank	Male	Non-shared words	Female	Non-shared words
1	Economy & Money	90/233 38.6%	Economy & Money	92/320 28.8%
2	Environment & Climate	70/275 25.5%	Sport & Physical Activities	70/403 17.4%
3	Sport & Physical Activities	46/291 15.8%	Environment & Climate	75/461 16.3%
4	Food & Drink	39/344 11.3%	Food & Drink	62/566 11%
5	Animals	34/343 9.9%	Animals	42/456 9.2%

Table 8.50*Non-shared Words in French by Gender*

9 th grade				
Rank	Male	Non-shared words	Female	Non-shared words
1	Economy & Money	45/94 47.9%	Economy & Money	47/236 17.5%
2	Sport & Physical Activities	29/132 22%	Environment & Climate	37/272 11.8%
3	Food & Drink	26/134 19.4%	Animals	25/292 11.2%
4	Environment & Climate	27/164 16.5%	Sport & Physical Activities	29/257 10.1%
5	Animals	20/142 14.1%	Food & Drink	28/269 8.7%
10 th grade				
Rank	Male	Non-shared words	Female	Non-shared words
1	Economy & Money	27/160 16.9%	Economy & Money	60/267 22.5%
2	Environment & Climate	17/114 14.9%	Environment & Climate	48/313 15.3%
3	Sport & Physical Activities	16/172 9.3%	Sport & Physical Activities	38/288 13.2%
4	Food & Drink	12/121 9.9%	Food & Drink	41/322 12.7%
5	Animals	9/126 7.1%	Animals	26/224 11.6%
11 th grade				
Rank	Male	Non-shared words	Female	Non-shared words
1	Environment & Climate	38/119 31.9%	Economy & Money	54/231 23.4%
2	Economy & Money	29/110 26.4%	Environment & Climate	56/316 17.7%
3	Food & Drink	28/116 24.1%	Food & Drink	50/290 17.2%
4	Sport & Physical Activities	18/148 12.2%	Sport & Physical Activities	33/218 15.1%
5	Animals	13/139 9.4%	Animals	23/259 8.9%

Regarding French, greater variety was found across grades. The greatest differences were found in 9th grade. Both genders produced the highest number of non-shared words in *Economy & Money*. However, while *Sport & Physical Activities* and *Food & Drink* came second and third for male students, they were fourth and fifth for female students, while *Environment & Climate* and *Animal* came second and third for the male students, they were fourth and fifth for female students. As in the case of English, male students were found to have a higher lexical sophistication than female students in terms of non-shared words, given that they had a higher percentage across all categories. The ranking for male and female students in 10th grade, on the other hand, was exactly the same, while in 11th only minor differences were found in *Environment & Climate* and *Economy & Money*, the former coming first for males and second for females and the latter coming first for females and second for males. In 11th grade, male students again produced a higher number of non-shared words in the majority of categories, with the exception of *Sport & Physical Activities*. However, contrary to the results for English and for 9th and 11th grade students in French, 10th grade female students produced a higher number of non-shared words than male students across all prompts. Thus, for French, it appears that 10th grade female students demonstrated a higher level of lexical sophistication with regard to non-shared words in the prompts at hand.

The final component of the qualitative analysis was an analysis of the number of infrequent words by male and female students in the production of each prompt for English (Table 8.51) and for French (Table 8.52). For English, the results were notably quite similar for male and female students in all grades, in particular for the general prompts *Animals* and *Food & Drink*. Regarding the content-specific prompts, the largest differences were seen at the K1-K5 band and off-list levels. In 9th grade, male students had a higher percentage of words at the K1-K5 band and a lower percentage of off-list words in the prompt *Sport & Physical Activities*, while the reverse was true for the prompts *Environment and Climate* and *Economy & Money*. Male students also produced a much higher percentage of words at the K6-K10 band in the prompt *Economy & Money*. In 10th grade, the main differences observed were in the two English content-relevant prompts, *Sport & Physical Activities* and *Economy & Money*. In both cases, female students produced a higher percentage of words at the K1-K5 band, while male students had a higher percentage of off-list words for *Sport & Physical Activities* and of words at the K6-K10 band for the prompt *Economy & Money*. Finally, in 11th grade, some minor differences were observed in the prompts *Environment and Climate* and *Economy &*

Money. For the former, female students produced a higher percentage of words at the K1-K5 band, whereas for the latter, male students had a higher percentage of words at the K6-K10 band and a lower percentage of off-list words. It should be noted that across all groups, female students very often produced more words than male students at the K1-K5 band, with the exception of the prompt *Animals* at all grades and the prompt *Sport & Physical Activities* in 9th grade. Notably these two prompts were the only two which revealed no statistically significant difference in 9th grade in the quantitative analysis, while the prompt *Animals* also revealed no statistically significant difference in 10th grade. It could thus be suggested that, while female students generally produce a higher number of words than their male peers, this may be accomplished by producing a greater amount of simple vocabulary from the K1-K5 band.

As in the case of English, results for French were in general quite similar for male and female students, with comparable findings across grades but often for different prompts. In 9th grade, female students produced a higher percentage of words at the K1-K5 band for the prompts *Food & Drink* and *Economy & Money* and at the K16-K20 and K21-25 bands for the prompt *Sport & Physical Activities*, while male students produced a much higher percentage of off-list words in these three prompts. A similar trend was found in 10th grade, where female students produced a higher percentage of words at the K1-K5 band for almost all prompts, with the exception of *Animals*, while male students often produced a much higher percentage of off-list words in these four prompts. Finally, in 11th grade, female students again produced a higher percentage of words at the K1-K5 band and male students produced a much higher percentage of off-list words, but in this case only in the prompt *Sport & Physical Activities*. Thus, it could again be argued that, although female students produce a quantitatively higher number of words than their male peers, this be done by relying on using simple vocabulary from the K1-K5 band.

Table 8.51*Frequency Distributions English by Gender*

9 th grade							
Prompt		K1-K5	K6-K10	K11-K15	K16-K20	K21-K25	Off-list
1	Male	78.3	14.7	2.5	2.6	0.4	1.48
	Female	74.7	18.6	2.3	2.3	0	2.11
2	Male	77.4	14.1	4.7	1.4	0	2.7
	Female	80.6	10.2	3.8	0.7	0	4.55
3	Male	71.6	14.2	1.7	0	0	12.44
	Female	65.6	14.1	2	0	0	18.26
4	Male	88.2	3.5	0.4	0	0	7.93
	Female	93	1.3	0.6	0	0	4.91
5	Male	75.6	17.8	0	0	0	6.5
	Female	85.6	5.7	0	0	0	8.8
10 th grade							
Prompt		K1-K5	K6-K10	K11-K15	K16-K20	K21-K25	Off-list
1	Male	78.9	15.8	2.8	0.4	0	2.26
	Female	77.1	16.8	3.1	1	0.2	1.66
2	Male	76.2	13.4	2.9	1.3	0	6.04
	Female	78.5	11.1	3.6	0.3	0	6.29
3	Male	59.5	12	2.6	0	0	25.89
	Female	71.8	10.1	1.3	0	0.4	16.41
4	Male	85.6	3.8	0.4	0	0.4	9.61
	Female	88.5	3.7	0.4	0	0.1	7.12
5	Male	78.7	8.6	1.1	0.6	0	10.92
	Female	86.1	3.7	0	0	0	10.3
11 th grade							
Prompt		K1-K5	K6-K10	K11-K15	K16-K20	K21-K25	Off-list
1	Male	79	16.6	1.8	0	0	2.62
	Female	76.7	17.1	1.9	1.2	0	2.85
2	Male	74.2	16.5	3.8	0.6	0	4.94
	Female	75.7	12	5.2	0.7	0.2	6.36
3	Male	65.3	10	2.7	0.7	0	21.31
	Female	69.9	9.9	0.9	0.2	0	18.86
4	Male	83.6	4.7	0.4	0	0.4	10.91
	Female	90	2.3	0.2	0	0.2	7.38
5	Male	80.3	11.5	2.5	0	0	5.58
	Female	81.2	5.9	0.3	0	0	12.5

Note. Prompt 1 = *Animals*, Prompt 2 = *Food & Drink*, Prompt 3 = *Sport and Physical Activities*, Prompt 4 = *Environment & Climate*, Prompt 5 = *Economy & Money*.

Table 8.52*Frequency Distributions French by Gender*

9 th grade							
Prompt		K1-K5	K6-K10	K11-K15	K16-K20	K21-K25	Off-list
1	Male	45.1	46.5	0	3.5	2.8	2.11
	Female	47.5	39.1	2.4	7.8	2.3	0.68
2	Male	56	19.4	7.4	1.5	4.4	10.45
	Female	63.8	17.8	10.4	1.5	3.3	2.97
3	Male	36.3	28.1	0	1.6	3.1	31.06
	Female	36.2	29.6	4.3	7.8	4.7	17.51
4	Male	89.1	4.2	3.6	0.6	0	2.44
	Female	86	8.9	1.1	1.5	0.4	2.21
5	Male	81.9	5.4	2.2	0	0	10.64
	Female	91.1	3.8	1.2	0.4	0	3.39
10 th grade							
Prompt		K1-K5	K6-K10	K11-K15	K16-K20	K21-K25	Off-list
1	Male	58.7	35.7	0.8	0.8	2.4	1.59
	Female	58.5	29.5	2.5	7.1	0.9	1.34
2	Male	56.2	12.4	8.2	4.1	4.1	14.88
	Female	51.5	23.6	14.3	1.2	2.9	6.21
3	Male	26.2	30.8	3.5	1.8	7.5	30.23
	Female	40.6	32.3	5.1	3.4	3.3	14.93
4	Male	83.3	7	0.9	2.7	0.9	5.26
	Female	90.1	4.8	3	0	0.3	1.92
5	Male	68.7	17.5	0.6	0	1.2	11.88
	Female	91.4	3.7	0.4	0	0	4.48
11 th grade							
Prompt		K1-K5	K6-K10	K11-K15	K16-K20	K21-K25	Off-list
1	Male	48.2	37.4	0.7	7.9	2.2	3.6
	Female	52.6	38.3	1.6	3.5	2.8	1.54
2	Male	53.4	18.1	11.2	2.6	6.9	7.76
	Female	56.6	18.3	7.5	1.7	2.8	13.1
3	Male	25.7	29.1	6.9	3.4	7.4	27.7
	Female	47.8	29.4	2.4	4.6	1.9	14.22
4	Male	83.3	4.9	0.8	0.8	0.8	9.24
	Female	80.6	6.9	2.2	0.9	0	9.18
5	Male	90.1	4.5	0	0	0.9	4.55
	Female	88.3	3.9	0.4	0.4	0.8	6.06

Note. Prompt 1 = *Animals*, Prompt 2 = *Food & Drink*, Prompt 3 = *Sport and Physical Activities*, Prompt 4 = *Environment & Climate*, Prompt 5 = *Economy & Money*.

In summary, analysis of the frequency of first word responses showed no differences between male and female students in the prompts *Animals*, *Food & Drink*, *Environment & Climate* in either language. However, for *Sport & Environment*, male students' most frequent first word in both languages was "football", while females' most frequent first word varied to a greater extent in language and grade. For *Economy & Money*, differences were again found, however, most responses tended to appear in the prompt name itself ("economy" / "money") or to be a type of currency ("euro" / "dollar"). Analysis of the most and least productive prompts showed a great deal of similarity between male and female students in English in all three grades and in French in 9th grade. However, the prompt *Sport & Physical Activities* ranked first for male students in both 10th and 11th grade, whereas for females it ranked third in 10th grade and last in 11th grade. The prompt *Environment & Climate* ranked second in 10th grade and first in 11th grade for female, whereas it ranked last in 10th grade and third in 11th grade for male students. Regarding lexical sophistication based on the non-shared words, clear similarity was found in the ranking of male and female students in each grade for English, however, male students notably had a higher percentage of non-shared words than their female peers in almost all cases, suggesting a higher level of lexical sophistication. Greater differences were found in the rankings across grades for French, and although 9th and 11th male students again had a higher percentage of non-shared words than their female peers in almost all cases, 10th grade female students produced a higher number of non-shared words than male students across all prompts. Finally, in the analysis of the lexical sophistication based on the number of infrequent words in the production of each prompt, results were generally quite similar for male and female students in all grades. One notable trend was that female students often produced a higher percentage of words at the K1-K5 band, while male students produced a higher percentage of off-list words. This suggests that, although female students produce a higher number of words than male students, male students produce vocabulary which is more lexically sophisticated.

8.3.3. Quantitative Differences in Gender and Motivation

Research question 3.3 asked whether there were quantitative differences between male and female students' language learning motivation in English and French at each testing period. As in the case of LA, two groups of analyses were carried out. The first investigated male and female students' motivation in English as compared to French at each testing period, that is, within gender and across languages (RQ3.3.1), while the

second addressed male participants' motivation as compared to female participants' motivation in English and in French at each testing period, that is, across gender and within languages (RQ3.3.2).

8.3.3.1. Quantitative Differences in Gender and Motivation across Languages.

Research question 3.1.1 asked whether there were quantitative differences in male and female participants' motivation in English as compared to French in each grade at each testing period. Table 8.53 provides the descriptive statistics for motivation for each language.

Table 8.53

Descriptive Statistics for Language Learning Motivation within Gender across Languages

9 th grade: Male								
English (<i>n</i> = 16)				Category	French (<i>n</i> = 18)			
Min	Max	<i>Mdn</i>	<i>SD</i>		Min	Max	<i>Mdn</i>	<i>SD</i>
1.80	4.80	3.80	.89	Ideal L2 Self	1.20	4.20	2.70	.98
1.57	3.71	2.57	.57	The "Ought to" Self	1.86	3.29	2.54	.42
2.20	5.00	4.40	.90	Language Anxiety	1.00	5.00	4.00	1.21
2.33	4.83	3.83	.88	Interest in FLs	1.00	4.33	2.75	.97
2.25	5.00	3.88	.78	L2 Self Confidence	1.00	4.50	3.00	.86
1.00	5.00	3.15	.08	Instrumentality: Prevention	1.20	4.60	2.30	.91
1.50	5.00	3.96	.89	Instrumentality: Promotion	1.00	4.83	2.83	.99
1.86	4.71	3.60	.84	Attitude towards Learning	1.25	4.00	2.63	.79
2.60	4.80	3.80	.57	Intended Learning Effort	1.00	3.80	3.00	.70
2.98	4.51	3.48	.51	Mean Motivation	1.77	3.48	2.89	.47
9 th grade: Female								
English (<i>n</i> = 23)				Category	French (<i>n</i> = 23)			
Min	Max	<i>Mdn</i>	<i>SD</i>		Min	Max	<i>Mdn</i>	<i>SD</i>
2.00	5.00	4.20	.76	Ideal L2 Self	1.80	5.00	2.80	.97
1.71	3.71	2.57	.60	The "Ought to" Self	1.14	3.57	2.57	.76
1.80	5.00	3.80	.91	Language Anxiety	1.20	5.00	3.60	1.04
2.83	5.00	3.83	.62	Interest in FLs	1.50	5.00	3.17	.88
2.25	5.00	3.75	.64	L2 Self Confidence	2.25	4.50	3.50	.57
3.20	5.00	4.20	.49	Instrumentality: Prevention	1.80	4.80	3.00	.71
2.33	5.00	4.20	.60	Instrumentality: Promotion	1.33	4.83	3.50	.85
1.50	4.69	3.36	.82	Attitude towards Learning	1.25	4.50	2.38	.84
2.20	5.00	4.00	.63	Intended Learning Effort	1.40	5.00	3.20	1.03
2.86	4.46	3.72	.45	Mean Motivation	2.13	4.29	2.99	.56

Table 8.53 (continued)

10 th grade: Male								
English (<i>n</i> = 15)					French (<i>n</i> = 15)			
Min	Max	<i>Mdn</i>	<i>SD</i>	Category	Min	Max	<i>Mdn</i>	<i>SD</i>
2.80	4.60	4.20	.51	Ideal L2 Self	1.80	4.40	3.40	.77
2.14	3.71	3.00	.55	The “Ought to” Self	2.00	3.71	3.00	.52
2.40	5.00	4.20	.89	Language Anxiety	2.80	5.00	4.00	.66
2.83	4.67	4.00	.58	Interest in FLs	2.17	4.50	3.33	.62
2.00	5.00	3.75	.84	L2 Self Confidence	2.25	4.50	3.50	.65
2.40	4.60	3.60	.61	Instrumentality: Prevention	2.20	4.20	3.40	.49
3.17	4.83	4.33	.42	Instrumentality: Promotion	1.17	4.83	3.67	.85
3.00	4.40	3.60	.47	Attitude towards Learning	1.38	4.00	3.25	.66
2.80	4.60	3.80	.50	Intended Learning Effort	2.40	4.00	3.20	.61
2.93	4.29	3.79	.40	Mean Motivation	2.21	3.99	3.46	.45
10 th grade: Female								
English (<i>n</i> = 26)					French (<i>n</i> = 26)			
Min	Max	<i>Mdn</i>	<i>SD</i>	Category	Min	Max	<i>Mdn</i>	<i>SD</i>
2.60	5.00	3.80	.70	Ideal L2 Self	1.00	4.40	2.90	.93
2.29	3.86	3.00	.41	The “Ought to” Self	1.57	3.57	2.57	.48
1.60	5.00	3.80	.91	Language Anxiety	1.20	5.00	3.30	.98
2.17	4.83	3.67	.61	Interest in FLs	1.67	4.17	3.42	.66
2.00	5.00	3.63	.80	L2 Self Confidence	1.00	5.00	3.25	.93
2.40	5.00	4.00	.66	Instrumentality: Prevention	1.20	4.00	2.90	.71
3.00	5.00	4.00	.51	Instrumentality: Promotion	1.67	4.67	3.50	.76
2.11	4.30	3.55	.55	Attitude towards Learning	1.00	3.75	2.63	.77
2.40	4.80	3.80	.57	Intended Learning Effort	1.20	4.40	2.90	.88
2.98	4.42	3.73	.42	Mean Motivation	1.61	4.04	3.02	.57
11 th grade: Male								
English (<i>n</i> = 19)					French (<i>n</i> = 15)			
Min	Max	<i>Mdn</i>	<i>SD</i>	Category	Min	Max	<i>Mdn</i>	<i>SD</i>
2.60	5.00	4.20	.71	Ideal L2 Self	1.60	4.20	3.20	.70
1.86	3.57	3.00	.52	The “Ought to” Self	1.14	3.14	2.71	.54
2.40	5.00	4.20	.89	Language Anxiety	2.60	5.00	4.00	.72
2.83	4.83	3.83	.59	Interest in FLs	2.25	4.33	3.33	.66
2.00	4.50	3.75	.68	L2 Self Confidence	2.25	4.50	3.25	.52
2.60	5.00	3.80	.65	Instrumentality: Prevention	1.40	4.00	3.00	.93
2.67	5.00	4.33	.65	Instrumentality: Promotion	1.67	4.83	3.50	.94
1.75	4.30	3.42	.69	Attitude towards Learning	1.63	4.00	3.13	.72
1.80	4.60	3.80	.72	Intended Learning Effort	1.60	4.00	3.20	.78
2.77	4.23	3.80	.43	Mean Motivation	2.15	3.91	3.16	.50

Table 8.53 (continued)

11 th grade: Female								
English (<i>n</i> = 21)				Category	French (<i>n</i> = 21)			
Min	Max	<i>Mdn</i>	<i>SD</i>		Min	Max	<i>Mdn</i>	<i>SD</i>
3.00	5.00	4.40	.67	Ideal L2 Self	1.20	4.40	2.80	1.02
1.71	4.43	3.00	.68	The “Ought to” Self	1.14	3.86	2.57	.64
2.00	5.00	4.00	.85	Language Anxiety	1.80	5.00	3.60	1.10
2.50	5.00	3.83	.67	Interest in FLs	1.83	4.50	3.50	.72
1.75	5.00	3.75	.92	L2 Self Confidence	1.25	5.00	3.25	.88
2.80	5.00	4.00	.61	Instrumentality: Prevention	1.00	4.40	2.20	.92
3.17	5.00	4.00	.49	Instrumentality: Promotion	1.83	5.00	3.00	.87
2.00	4.60	3.10	.63	Attitude towards Learning	1.38	4.00	3.00	.56
2.40	4.80	4.00	.53	Intended Learning Effort	1.00	4.00	2.80	.80
3.13	4.59	3.78	.39	Mean Motivation	1.85	3.93	3.10	.54

As shown, both male and female participants in each grade reported higher motivation in English than in French overall and in almost all categories, with the exception of the category *The “Ought to” Self* in 9th grade female participants and 10th grade male participants, where no difference was observed in the medians.

In order to determine whether these differences were statistically significant, and in particular to pinpoint whether there were any differences between male and female learners in their results, Wilcoxon signed-rank tests were conducted, comparing the results of the nine categories of the MFQ, as well as the overall MFQ mean of the participants in English and French (Table 8.54).

Table 8.54

Differences in Language Learning Motivation within Gender across Languages

9 th Grade: Male						
Category	<i>Mdn</i>		<i>SD</i>		<i>z</i>	<i>p</i>
	En	Fr	En	Fr		
Ideal L2 Self	3.80	2.70	.89	.98	-2.73	.006
The “Ought to” Self	2.57	2.54	.57	.42	-1.32	.187
Language Anxiety	4.40	4.00	.90	1.21	-2.31	.021
Interest in Foreign Languages	3.83	2.75	.88	.97	-2.65	.008
L2 Self Confidence	3.88	3.00	.78	.86	-2.13	.033
Instrumentality: Prevention	3.15	2.30	.08	.91	-3.06	.002
Instrumentality: Promotion	3.96	2.83	.89	.99	-3.03	.002
Attitude towards Learning	3.60	2.63	.84	.79	-2.73	.006
Intended Learning Effort	3.80	3.00	.57	.70	-2.99	.003
Mean Motivation	3.48	2.89	.51	.47	-3.41	.001

Table 8.54 (continued)

9 th Grade: Female						
Category	<i>Mdn</i>		<i>SD</i>		<i>z</i>	<i>p</i>
	En	Fr	En	Fr		
Ideal L2 Self	4.20	2.80	.76	.97	-4.11	<.001
The “Ought to” Self	2.57	2.57	.60	.76	-1.22	.223
Language Anxiety	3.80	3.60	.91	1.04	-1.42	.156
Interest in Foreign Languages	3.83	3.17	.62	.88	-3.35	.001
L2 Self Confidence	3.75	3.50	.64	.57	-2.18	.029
Instrumentality: Prevention	4.20	3.00	.49	.71	-4.05	<.001
Instrumentality: Promotion	4.20	3.50	.60	.85	-3.92	<.001
Attitude towards Learning	3.36	2.38	.82	.84	-3.32	.001
Intended Learning Effort	4.00	3.20	.63	1.03	-4.02	<.001
Mean Motivation	3.72	2.99	.45	.56	-4.17	<.001
10 th Grade: Male						
Category	<i>Mdn</i>		<i>SD</i>		<i>z</i>	<i>p</i>
	En	Fr	En	Fr		
Ideal L2 Self	4.20	3.40	.51	.77	-3.14	.002
The “Ought to” Self	3.00	3.00	.55	.52	-2.26	.024
Language Anxiety	4.20	4.00	.89	.66	-1.02	.306
Interest in Foreign Languages	4.00	3.33	.58	.62	-2.16	.031
L2 Self Confidence	3.75	3.50	.84	.65	-0.60	.548
Instrumentality: Prevention	3.60	3.40	.61	.49	-1.26	.208
Instrumentality: Promotion	4.33	3.67	.42	.85	-2.99	.003
Attitude towards Learning	3.60	3.25	.47	.66	-2.61	.009
Intended Learning Effort	3.80	3.20	.50	.61	-2.67	.007
Mean Motivation	3.79	3.46	.40	.45	-2.67	.008
10 th Grade: Female						
Category	<i>Mdn</i>		<i>SD</i>		<i>z</i>	<i>p</i>
	En	Fr	En	Fr		
Ideal L2 Self	3.80	2.90	.70	.93	-4.15	<.001
The “Ought to” Self	3.00	2.57	.41	.48	-4.00	<.001
Language Anxiety	3.80	3.30	.91	.98	-2.15	.032
Interest in Foreign Languages	3.67	3.42	.61	.66	-2.34	.019
L2 Self Confidence	3.63	3.25	.80	.93	-1.69	.090
Instrumentality: Prevention	4.00	2.90	.66	.71	-4.38	<.001
Instrumentality: Promotion	4.00	3.50	.51	.76	-4.13	<.001
Attitude towards Learning	3.55	2.63	.55	.77	-3.94	<.001
Intended Learning Effort	3.80	2.90	.57	.88	-4.04	<.001
Mean Motivation	3.73	3.02	.42	.57	-4.13	<.001

Table 8.54 (continued)

11 th Grade: Male						
Category	Mdn		SD		z	p
	En	Fr	En	Fr		
Ideal L2 Self	4.20	3.20	.71	.70	-3.21	.001
The “Ought to” Self	3.00	2.71	.52	.54	-2.91	.004
Language Anxiety	4.20	4.00	.89	.72	-1.27	.205
Interest in Foreign Languages	3.83	3.33	.59	.66	-1.74	.082
L2 Self Confidence	3.75	3.25	.68	.52	-0.71	.477
Instrumentality: Prevention	3.80	3.00	.65	.93	-3.05	.002
Instrumentality: Promotion	4.33	3.50	.65	.94	-3.30	.001
Attitude towards Learning	3.42	3.13	.69	.72	-2.16	.031
Intended Learning Effort	3.80	3.20	.72	.78	-2.46	.014
Mean Motivation	3.80	3.16	.43	.50	-3.07	.002
11 th Grade: Female						
Category	Mdn		SD		z	p
	En	Fr	En	Fr		
Ideal L2 Self	4.40	2.80	.67	1.02	-3.39	.001
The “Ought to” Self	3.00	2.57	.68	.64	-2.02	.044
Language Anxiety	4.00	3.60	.85	1.10	-2.00	.046
Interest in Foreign Languages	3.83	3.50	.67	.72	-1.87	.061
L2 Self Confidence	3.75	3.25	.92	.88	-1.43	.152
Instrumentality: Prevention	4.00	2.20	.61	.92	-3.88	<.001
Instrumentality: Promotion	4.00	3.00	.49	.87	-3.62	<.001
Attitude towards Learning	3.10	3.00	.63	.56	-0.73	.465
Intended Learning Effort	4.00	2.80	.53	.80	-3.81	<.001
Mean Motivation	3.78	3.10	.39	.54	-3.67	<.001

Results showed that there were statistically significant differences between both male and female participants' motivation towards English and French in all grades for overall motivation, and in almost all individual categories, with the following exceptions: *The “Ought to” Self* in 9th grade male participants; *The “Ought to” Self* and *Language Anxiety* in female 9th grade participants; *Language Anxiety*, *L2 Self Confidence*, and *Instrumentality: Prevention* in 10th grade male participants, *L2 Self Confidence* in 10th grade female participants; *Language Anxiety*, *Interest in Foreign Languages* and *L2 Self Confidence* in 11th grade male participants; and *Interest in Foreign Languages*, *L2 Self Confidence* and *Attitude Towards Learning* in 11th grade female participants. What is of most interest here are the differences between male and female students, which were observed in the case of *Language Anxiety* in 9th grade, *Language Anxiety*, and *Instrumentality: Prevention* in 10th grade, *Language Anxiety* and *Attitude Towards Learning* in 11th grade. In 9th grade, statistically significant differences were found for

Language Anxiety for male learners but not for female learners. This indicates that while male students report lower anxiety for English than for French, female learners report similar levels of anxiety for both languages. In 10th grade, statistically significant differences were found for *Language Anxiety* and *Instrumentality: Prevention* for female learners but not for male learners. This suggests that while female students report lower anxiety for English than for French, male learners report similar levels of anxiety for both languages. In addition, female students report that not having English will prevent their future success to a greater extent than not having French, whereas for male learners there is no difference between the two languages. Finally, in 11th grade, statistically significant differences were again found for *Language Anxiety* for female learners but not for male learners, and for *Attitude Towards Learning* for male learners but not for female learners. This indicates that while female students report lower anxiety and for English than for French, male learners report similar levels of anxiety for both languages. On the other hand, while male participants expressed a more positive attitude towards learning English than French, female participants' attitude towards the two languages showed no significant difference. The results are summarised in Table 8.55 and presented visually in Figure 8.17.

Table 8.55

Summary of Quantitative Differences in Gender and Motivation across Languages

		Male	Female
9 th grade	<i>Language Anxiety</i>	✓	
10 th grade	<i>Language Anxiety</i>		✓
	<i>Instrumentality: Prevention</i>		✓
11 th grade	<i>Language Anxiety</i>		✓
	<i>Attitude Towards Learning</i>	✓	

Note. ✓ = statistically significant difference between English and French

Figure 8.17

Differences in Language Learning Motivation within Gender across Languages

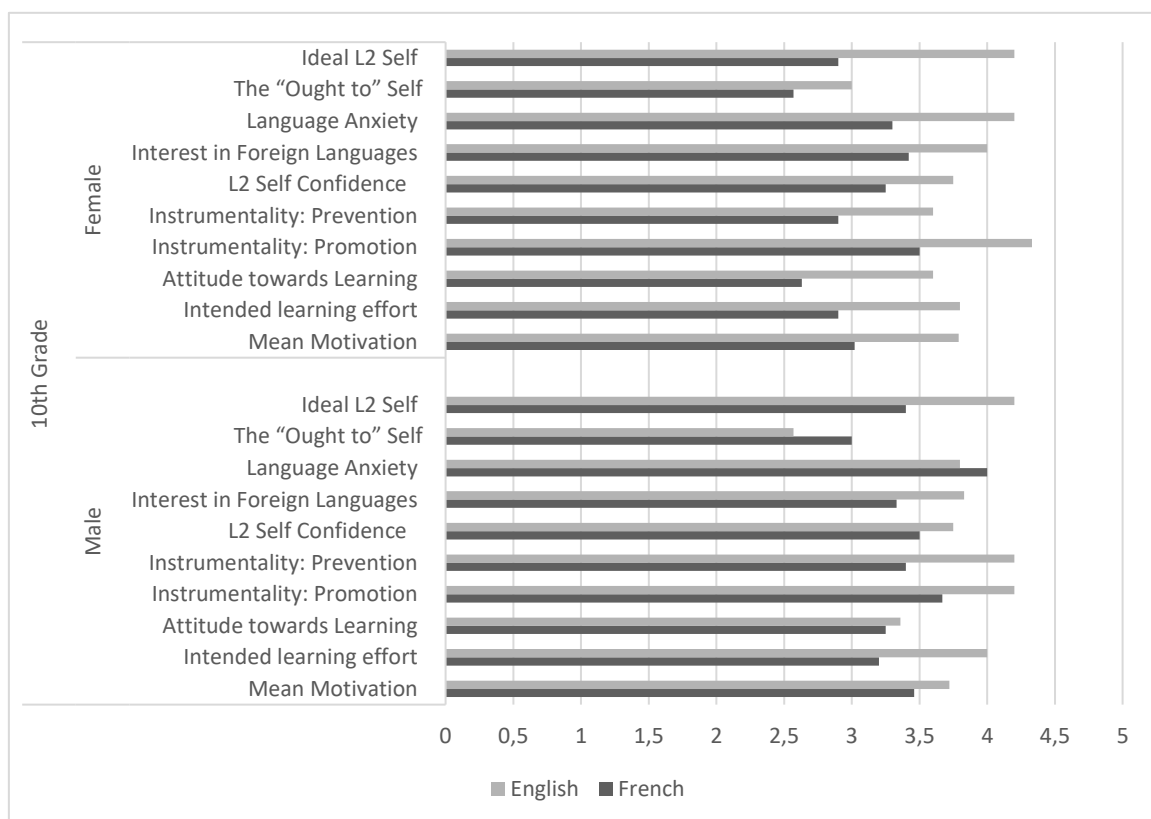
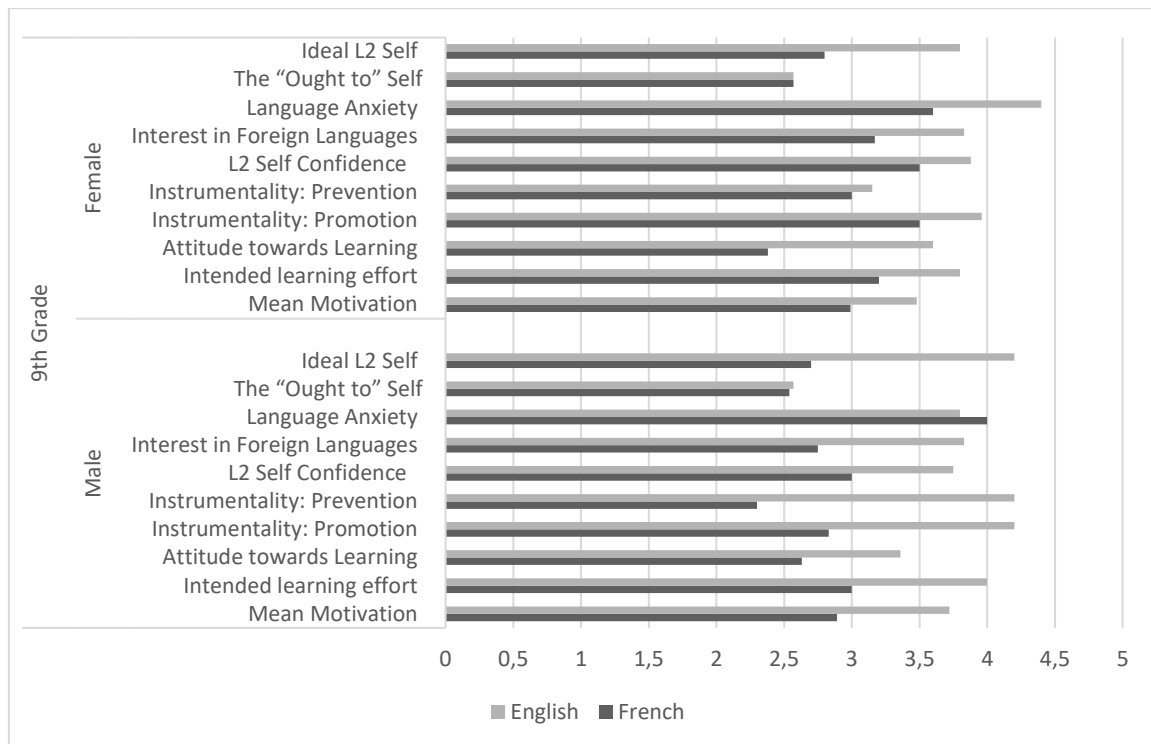
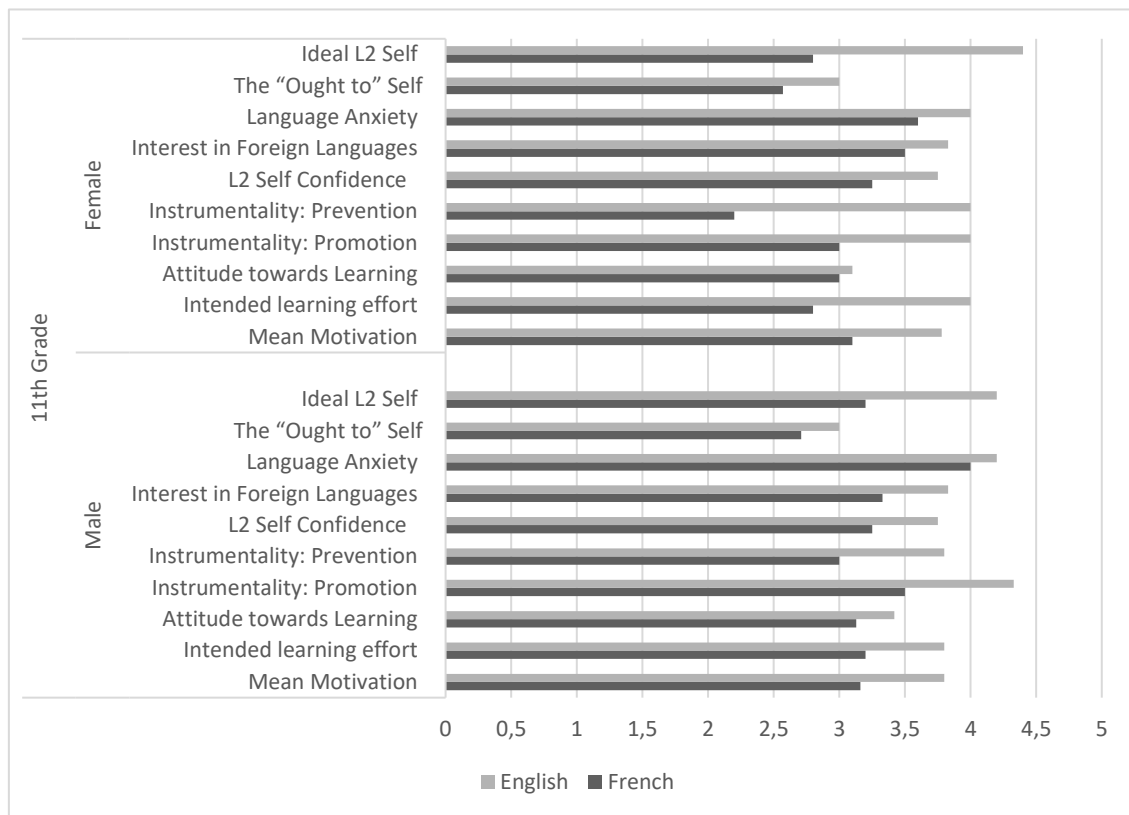


Figure 8.17 (continued)

Differences in Language Learning Motivation within Gender across Languages



In order to determine whether there were cross-sectional and longitudinal differences in the participants' language learning motivation in relation to gender and language, Mann-Whiney U tests were carried out to compare students in 9th grade to students in 10th grade and Wilcoxon signed-rank tests were carried out to compare students in 10th grade to 11th grade. The results were then analysed to compare the results of male students in English versus French, and of female students in English versus French.

Firstly, regarding the cross-sectional analysis, results revealed that in the case of male students there were no statistically significant differences between 9th grade and 10th grade students for English language learning motivation, while for French language learning motivation statistically significant differences were found between 9th grade and 10th grade students in the participants' mean French language learning motivation as well as a number of categories: *The "Ought to" Self*, *Interest in Foreign Languages*, *Instrumentality: Prevention* and *Instrumentality: Promotion*. In all cases, means were higher for 10th grade students than for 9th grade students (Table 8.56).

Table 8.56

Cross-Sectional Differences in Language Learning Motivation by Male Participants in English and French

9 th → 10 th grade: English						
	<i>Mdn</i>		<i>SD</i>		<i>z</i>	<i>p</i>
	9 th	10 th	9 th	10 th		
Ideal L2 Self	3.80	4.20	.888	.506	-1.07	.283
The “Ought to” Self	2.57	3.00	.566	.551	-1.35	.176
Language Anxiety	4.40	4.20	.905	.891	-0.54	.590
Interest in Foreign Languages	3.83	4.00	.885	.580	-0.24	.811
L2 Self Confidence	3.88	3.75	.778	.837	-0.66	.510
Instrumentality: Prevention	3.15	3.60	.084	.612	-1.23	.219
Instrumentality: Promotion	3.96	4.33	.886	.425	-0.72	.474
Attitude towards Learning	3.60	3.60	.837	.473	-0.71	.476
Intended Learning Effort	3.80	3.80	.575	.504	-0.34	.734
Mean English Motivation	3.47	3.79	.512	.402	-0.91	.363
9 th → 10 th grade: French						
	<i>Mdn</i>		<i>SD</i>		<i>z</i>	<i>p</i>
	9 th	10 th	9 th	10 th		
Ideal L2 Self	2.70	3.40	.979	.767	-1.96	.050
The “Ought to” Self	2.54	3.00	.416	.519	-1.98	.048
Language Anxiety	4.00	4.00	1.21	.656	-0.51	.611
Interest in Foreign Languages	2.75	3.33	.967	.620	-2.14	.032
L2 Self Confidence	3.00	3.50	.857	.651	-1.53	.127
Instrumentality: Prevention	2.30	3.40	.915	.493	-3.08	.002
Instrumentality: Promotion	2.83	3.67	.992	.851	-2.65	.008
Attitude towards Learning	2.63	3.25	.788	.662	-1.98	.048
Intended Learning Effort	3.00	3.20	.702	.609	-0.24	.813
Mean French Motivation	2.89	3.46	.465	.447	-3.29	.001

On the other hand, in the case of female students, a statistically significant difference between 9th grade and 10th grade students for English language learning motivation was found only in the case of *The “Ought to” Self*, which was higher in 10th grade, while for French language learning motivation no statistically significant differences were found between 9th grade and 10th grade students (Table 8.57). These results indicate that there are clear differences between language learning motivation towards English and French depending on gender. Male students’ motivation towards English is relatively similar in 9th and 10th grade, while the older students clearly show higher motivation towards French than their younger peers. In particular, they report higher motivation due to external sources, greater interest in learning French, see French as more important in promoting their future success, and see not having French as more

preventative to their future success. However, much fewer differences arise in motivation towards each language for female students: motivation towards both English and French is relatively similar in 9th and 10th grade, with just one exception: for English, 10th grade female students report higher motivation due to external sources than their younger female peers.

Table 8.57

Cross-Sectional Differences in Language Learning Motivation by Female Participants in English and French

9 th → 10 th grade: English						
	<i>Mdn</i>		<i>SD</i>		<i>z</i>	<i>p</i>
	9 th	10 th	9 th	10 th		
Ideal L2 Self	4.20	3.80	.761	.699	-0.62	.537
The “Ought to” Self	2.57	3.00	.602	.409	-2.43	.015
Language Anxiety	3.80	3.80	.914	.913	-0.10	.920
Interest in Foreign Languages	3.83	3.67	.617	.607	-1.32	.186
L2 Self Confidence	3.75	3.63	.638	.800	-1.07	.285
Instrumentality: Prevention	4.20	4.00	.486	.662	-0.04	.968
Instrumentality: Promotion	4.20	4.00	.605	.512	-1.48	.138
Attitude towards Learning	3.36	3.55	.818	.547	-0.52	.602
Intended Learning Effort	4.00	3.80	.625	.569	-1.84	.066
Mean English Motivation	3.72	3.73	.445	.416	-0.58	.561
9 th → 10 th grade: French						
	<i>Mdn</i>		<i>SD</i>		<i>z</i>	<i>p</i>
	9 th	10 th	9 th	10 th		
Ideal L2 Self	2.80	2.90	.969	.935	-0.08	.936
The “Ought to” Self	2.57	2.57	.764	.479	-0.62	.534
Language Anxiety	3.60	3.30	1.04	.984	-0.60	.547
Interest in Foreign Languages	3.17	3.42	.878	.657	-0.22	.825
L2 Self Confidence	3.50	3.25	.569	.932	-0.94	.349
Instrumentality: Prevention	3.00	2.90	.707	.709	-0.95	.344
Instrumentality: Promotion	3.50	3.50	.852	.759	-0.44	.658
Attitude towards Learning	2.38	2.63	.838	.768	-0.04	.968
Intended Learning Effort	3.20	2.90	1.03	.875	-0.91	.361
Mean French Motivation	2.99	3.02	.558	.572	-0.18	.857

Secondly, regarding the longitudinal analysis, results revealed that in the case of male students there were statistically significant differences from 10th to 11th grade in the participants’ overall motivation and the category *The “Ought to” Self* both for English language learning motivation and for French language learning motivation (Table 8.58).

In all cases, means were lower in 11th grade than in 10th grade, suggesting a decrease in language learning motivation, regardless of the language at hand.

Table 8.58

Longitudinal Differences in Language Learning Motivation by Male Participants in English and French

10 th → 11 th grade: English						
	<i>Mdn</i>		<i>SD</i>		<i>Z</i>	<i>p</i>
	10 th	11 th	10 th	11 th		
Ideal L2 Self	4.20	4.20	.506	.715	-.444	.657
The “Ought to” Self	3.00	3.00	.551	.518	-2.14	.032
Language Anxiety	4.20	4.20	.891	.888	-.079	.937
Interest in Foreign Languages	4.00	3.83	.580	.591	-1.77	.076
L2 Self Confidence	3.75	3.75	.837	.680	-.694	.487
Instrumentality: Prevention	3.60	3.80	.612	.648	-.396	.692
Instrumentality: Promotion	4.33	4.33	.425	.651	-.599	.549
Attitude towards Learning	3.60	3.42	.473	.689	-1.85	.064
Intended Learning Effort	3.80	3.80	.504	.723	-1.91	.056
Mean English Motivation	3.79	3.80	.402	.435	-2.04	.041
10 th → 11 th grade: French						
	<i>Mdn</i>		<i>SD</i>		<i>Z</i>	<i>p</i>
	10 th	11 th	10 th	11 th		
Ideal L2 Self	3.40	3.20	.767	.705	-1.59	.111
The “Ought to” Self	3.00	2.71	.519	.544	-2.09	.036
Language Anxiety	4.00	4.00	.656	.718	-1.20	.227
Interest in Foreign Languages	3.33	3.33	.620	.661	-1.06	.288
L2 Self Confidence	3.50	3.25	.651	.516	-.632	.528
Instrumentality: Prevention	3.40	3.00	.493	.934	-1.96	.050
Instrumentality: Promotion	3.67	3.50	.851	.942	-1.61	.107
Attitude towards Learning	3.25	3.13	.662	.725	-1.72	.085
Intended Learning Effort	3.20	3.20	.609	.778	-1.01	.311
Mean French Motivation	3.46	3.16	.447	.502	-2.04	.041

On the other hand, in the case of female students a statistically significant difference was found between students in 10th and 11th for English language learning motivation only in the case of *Attitude towards Learning*, which was again lower in 11th grade, while for French language learning motivation no statistically significant differences were found from 10th to 11th grade (Table 8.59). These results again suggest that there are differences between language learning motivation towards English and French depending on gender. Male students’ language learning motivation from 10th to 11th grade was the same in each language: in both English and French, 11th grade students

reported lower motivation overall and lower motivation due to external sources. On the other hand, a slight difference was found between the languages for female learners, who reported a less positive attitude towards learning English in 11th grade than they had in 10th grade, while the same was not found in French.

Table 8.59

Longitudinal Differences in Language Learning Motivation by Female Participants in English and French

10 th → 11 th grade: English						
	<i>Mdn</i>		<i>SD</i>		<i>Z</i>	<i>p</i>
	10 th	11 th	10 th	11 th		
Ideal L2 Self	3.80	4.40	.699	.666	-1.00	.316
The “Ought to” Self	3.00	3.00	.409	.684	-.029	.977
Language Anxiety	3.80	4.00	.913	.854	-.380	.704
Interest in Foreign Languages	3.67	3.83	.607	.667	-.168	.867
L2 Self Confidence	3.63	3.75	.800	.917	-.493	.622
Instrumentality: Prevention	4.00	4.00	.662	.611	-.625	.532
Instrumentality: Promotion	4.00	4.00	.512	.493	-.190	.849
Attitude towards Learning	3.55	3.10	.547	.628	-2.74	.006
Intended Learning Effort	3.80	4.00	.569	.527	-.211	.833
Mean English Motivation	3.73	3.78	.416	.389	-.327	.744
10 th → 11 th grade: French						
	<i>Mdn</i>		<i>SD</i>		<i>Z</i>	<i>p</i>
	10 th	11 th	10 th	11 th		
Ideal L2 Self	2.90	2.80	.935	1.01	-.052	.959
The “Ought to” Self	2.57	2.57	.479	.644	-.591	.555
Language Anxiety	3.30	3.60	.984	1.10	-1.27	.204
Interest in Foreign Languages	3.42	3.50	.657	.717	-.119	.905
L2 Self Confidence	3.25	3.25	.932	.879	-1.06	.289
Instrumentality: Prevention	2.90	2.20	.709	.924	-1.67	.095
Instrumentality: Promotion	3.50	3.00	.759	.869	-1.83	.066
Attitude towards Learning	2.63	3.00	.768	.563	-1.19	.231
Intended Learning Effort	2.90	2.80	.875	.798	-1.62	.104
Mean French Motivation	3.02	3.10	.572	.544	-.283	.777

8.3.3.2. Quantitative Differences in Gender and Motivation within Languages.

Research question 3.1.2 asked whether there were quantitative differences in the male participants' motivation as compared to female participants' motivation in English and French in each grade at each testing period. In order to analyse this, results of each language learning motivation questionnaire were analysed in turn, comparing male and female students' first in English and then in French.

Regarding English, the descriptive statistics for the language learning motivation in each language, as shown in Table 8.60, indicated that in 9th grade, male students reported lower *Language Anxiety*, higher *L2 Self Confidence* and a better *Attitude towards Learning*, while female students reported higher motivation in terms of the *Ideal L2 Self*, *Instrumentality: Prevention*, *Instrumentality: Promotion*, *Intended Learning Effort* and overall motivation. No differences were observed in terms of *The "Ought to" Self* or *Interest in Foreign Languages*. In 10th grade, male students reported higher motivation than females in almost all motivation categories as well as overall motivation, with the exception of *Instrumentality: Prevention*, where females reported higher motivation, and *The "Ought to" Self* and *Intended learning effort*, where no differences were observed. In 11th grade, male students reported lower *Language Anxiety*, a better *Attitude towards Learning*, and higher *Instrumentality: Promotion* and overall motivation, while female students reported higher motivation in terms of the *Ideal L2 Self*, *Instrumentality: Prevention*, *Intended learning effort*. No differences were observed in terms of *The "Ought to" Self*, *Interest in Foreign Languages* or *L2 Self Confidence*. In order to determine whether these differences were statistically significant, Mann-Whiney U tests were conducted, comparing the results of the nine categories of the MFQ, as well as the overall motivation mean of the participants (Table 8.61). Results showed a statistically significant difference in just one of the categories in 9th grade, but not in 10th or 11th grade. In 9th grade, male students ($Mdn = 3.15$, $SD = 1.08$) reported a lower score for *Instrumentality: Prevention* than female students ($Mdn = 4.20$, $SD = 0.48$); $z = -2.69$, $p = .007$. This suggests that the female students in 9th grade see not having English as preventing their future success to a greater degree than the male students. There was no statistically significant difference between the male and female students in the overall MFQ score or any other categories at any level, indicating that both male and female students are similarly motivated towards learning English. These results are presented visually in Figure 8.18.

Table 8.60*Descriptive Statistics for Language Learning Motivation by Gender in English*

9 th grade								
Male (<i>n</i> = 16)					Female (<i>n</i> = 23)			
Min	Max	<i>Mdn</i>	<i>SD</i>	Category	Min	Max	<i>Mdn</i>	<i>SD</i>
1.80	4.80	3.80	.89	Ideal L2 Self	2.00	5.00	4.20	.76
1.57	3.71	2.57	.57	The “Ought to” Self	1.71	3.71	2.57	.60
2.20	5.00	4.40	.90	Language Anxiety	1.80	5.00	3.80	.91
2.33	4.83	3.83	.88	Interest in FLs	2.83	5.00	3.83	.62
2.25	5.00	3.88	.78	L2 Self Confidence	2.25	5.00	3.75	.64
1.00	5.00	3.15	1.08	Instrumentality: Prevention	3.20	5.00	4.20	.49
1.50	5.00	3.96	.89	Instrumentality: Promotion	2.33	5.00	4.20	.60
1.86	4.71	3.60	.84	Attitude towards Learning	1.50	4.69	3.36	.82
2.60	4.80	3.80	.57	Intended Learning Effort	2.20	5.00	4.00	.63
2.98	4.51	3.48	.51	Mean Motivation	2.86	4.46	3.72	.45
10 th grade								
Male (<i>n</i> = 15)					Female (<i>n</i> = 26)			
Min	Max	<i>Mdn</i>	<i>SD</i>	Category	Min	Max	<i>Mdn</i>	<i>SD</i>
2.80	4.60	4.20	.51	Ideal L2 Self	2.60	5.00	3.80	.70
2.14	3.71	3.00	.55	The “Ought to” Self	2.29	3.86	3.00	.41
2.40	5.00	4.20	.89	Language Anxiety	1.60	5.00	3.80	.91
2.83	4.67	4.00	.58	Interest in FLs	2.17	4.83	3.67	.61
2.00	5.00	3.75	.84	L2 Self Confidence	2.00	5.00	3.63	.80
2.40	4.60	3.60	.61	Instrumentality: Prevention	2.40	5.00	4.00	.66
3.17	4.83	4.33	.42	Instrumentality: Promotion	3.00	5.00	4.00	.51
3.00	4.40	3.60	.47	Attitude towards Learning	2.11	4.30	3.55	.55
2.80	4.60	3.80	.50	Intended Learning Effort	2.40	4.80	3.80	.57
2.93	4.29	3.79	.40	Mean Motivation	2.98	4.42	3.73	.42
11 th grade								
Male (<i>n</i> = 19)					Female (<i>n</i> = 21)			
Min	Max	<i>Mdn</i>	<i>SD</i>	Category	Min	Max	<i>Mdn</i>	<i>SD</i>
2.60	5.00	4.20	.71	Ideal L2 Self	3.00	5.00	4.40	.67
1.86	3.57	3.00	.52	The “Ought to” Self	1.71	4.43	3.00	.68
2.40	5.00	4.20	.89	Language Anxiety	2.00	5.00	4.00	.85
2.83	4.83	3.83	.59	Interest in FLs	2.50	5.00	3.83	.67
2.00	4.50	3.75	.68	L2 Self Confidence	1.75	5.00	3.75	.92
2.60	5.00	3.80	.65	Instrumentality: Prevention	2.80	5.00	4.00	.61
2.67	5.00	4.33	.65	Instrumentality: Promotion	3.17	5.00	4.00	.49
1.75	4.30	3.42	.69	Attitude towards Learning	2.00	4.60	3.10	.63
1.80	4.60	3.80	.72	Intended Learning Effort	2.40	4.80	4.00	.53
2.77	4.23	3.80	.43	Mean Motivation	3.13	4.59	3.78	.39

Table 8.61*Differences in Language Learning Motivation by Gender in English*

9 th Grade						
Category	Mdn		SD		z	p
	Male	Female	Male	Female		
Ideal L2 Self	3.80	4.20	.89	.76	-1.85	.065
The “Ought to” Self	2.57	2.57	.57	.60	-.33	.742
Language Anxiety	4.40	3.80	.90	.91	-1.47	.142
Interest in Foreign Languages	3.83	3.83	.88	.62	-.78	.437
L2 Self Confidence	3.88	3.75	.78	.64	-.49	.624
Instrumentality: Prevention	3.15	4.20	1.08	.49	-2.69	.007
Instrumentality: Promotion	3.96	4.20	.89	.60	-1.36	.173
Attitude towards Learning	3.60	3.36	.84	.82	-.54	.587
Intended Learning Effort	3.80	4.00	.57	.63	-1.87	.061
Mean English Motivation	3.48	3.72	.51	.45	-1.20	.230
10 th Grade						
Category	Mdn		SD		z	p
	Male	Female	Male	Female		
Ideal L2 Self	4.20	4.20	.51	.76	-.10	.924
The “Ought to” Self	3.00	2.57	.55	.60	-.03	.978
Language Anxiety	4.20	3.80	.89	.91	-1.03	.302
Interest in Foreign Languages	4.00	3.83	.58	.62	-1.06	.289
L2 Self Confidence	3.75	3.75	.84	.64	-.53	.596
Instrumentality: Prevention	3.60	4.20	.61	.49	-1.91	.056
Instrumentality: Promotion	4.33	4.20	.42	.60	-.85	.397
Attitude towards Learning	3.60	3.36	.47	.82	-.98	.328
Intended Learning Effort	3.80	4.00	.50	.63	-.72	.470
Mean English Motivation	3.79	3.72	.40	.45	-.38	.705
11 th Grade						
Category	Mdn		SD		z	p
	Male	Female	Male	Female		
Ideal L2 Self	4.20	4.40	.71	.67	-.95	.341
The “Ought to” Self	3.00	3.00	.52	.68	-.01	.989
Language Anxiety	4.20	4.00	.89	.85	-.07	.946
Interest in Foreign Languages	3.83	3.83	.59	.67	-.42	.673
L2 Self Confidence	3.75	3.75	.68	.92	-.82	.414
Instrumentality: Prevention	3.80	4.00	.65	.61	-1.39	.165
Instrumentality: Promotion	4.33	4.00	.65	.49	-.59	.557
Attitude towards Learning	3.42	3.10	.69	.63	-1.19	.233
Intended Learning Effort	3.80	4.00	.72	.53	-1.45	.147
Mean English Motivation	3.80	3.78	.43	.39	-.46	.645

Figure 8.18

Differences in Language Learning Motivation by Gender in English



While there were no significant differences in the other 8 categories, analyses were also run on the individual questions in order to reveal any minor differences that may be found (Table 8.62). In 9th, 10th and 11th grade, 4, 7 and 5 out of the fifty-five individual questions, respectively, yielded statistically significant differences between the groups, with just one question in common across all grades and one question in common in 9th and 10th grade. Although these results should be interpreted with caution, given that the overall category did not always show statistically significant differences, they provide some interesting findings which may reveal underlying differences between the groups in the present study.

Table 8.62*Statistically Significant Differences between Groups in Individual Questions in English*

9 th Grade						
Category	Mdn		SD		z	p
	Male	Fem	Male	Fem		
I can't imagine my future without English.	2.50	4.00	1.53	1.24	-2.01	.045
Given the economic situation in Spain, I'll need English to work abroad.	4.00	5.00	1.55	0.67	-1.97	.049
I don't want to fail with English because my professional future depends on it.	3.00	4.00	1.22	0.83	-2.13	.033
Learning biology and geology in English is interesting.	4.00	3.00	1.65	1.22	-2.18	.029
10 th Grade						
Category	Mdn		SD		z	p
	Male	Fem	Male	Fem		
I can't imagine my future without English.	3.00	4.00	.83	1.02	-2.06	.039
All my friends talk about the importance of learning English.	3.00	3.00	1.19	.93	-1.97	.049
I'm really curious about English structure and vocabulary.	4.00	3.00	.88	1.05	-3.42	.001
I don't want to fail with English because my professional future depends on it.	4.00	4.00	.99	.82	-2.26	.024
Learning English is important to me because I want to continue studying abroad.	4.00	3.00	.80	1.18	-2.79	.005
Learning economics in English is interesting.	5.00	2.00	.89	1.27	-2.38	.017
My experience in English class has always been positive.	4.00	4.00	.65	1.06	-1.98	.047

Table 8.62 (continued)

11 th Grade						
Category	<i>Mdn</i>		<i>SD</i>		<i>z</i>	<i>p</i>
	Male	Fem	Male	Fem		
I can't imagine my future without English.	3.00	5.00	1.28	.86	-2.77	.006
If I don't learn English, I won't be able to work as what I want.	3.00	4.00	1.23	1.05	-2.04	.042
Learning PE in English in secondary school was interesting.	3.00	2.00	1.29	1.26	-2.07	.038
I enjoyed PE classes in English in secondary school.	4.00	3.00	1.10	1.22	-1.98	.048
I think I'm doing all I can to learn English.	3.00	4.00	1.21	.89	-2.94	.003

In 9th grade the results suggested that the female group had a higher score than the male group with regard to three specific statements: “I can't imagine my future without English”, “Given the economic situation in Spain, I'll need English to work abroad”, and “I don't want to fail with English because my professional future depends on it”. The male group had a higher score than the female group with regard to just one specific statement: “Learning biology and geology in English is interesting”. In 10th grade, the results indicated that the female group had a higher score than the male group with regard to three specific statements: “I can't imagine my future without English”, “All my friends talk about the importance of learning English” and “I don't want to fail with English because my professional future depends on it”. They also show that the male group had a higher score than the female group with regard to four specific statements: “I'm really curious about English structure and vocabulary”, “Learning English is important to me because I want to continue studying abroad”, “Learning economics in English is interesting” and “My experience in English class has always been positive”. Finally, in 11th grade the results revealed that the female group had a higher score than the male group with regard to three specific statements: “I can't imagine my future without English”, “If I don't learn English, I won't be able to work as what I want” and “I think I'm doing all I can to learn English”. The male group had a higher score than the female group with regard to two specific statements: “Learning PE in English in secondary school was interesting” and “I enjoyed PE classes in English in secondary school”. As noted above, these results should be interpreted with caution, given that the overall category did not always show statistically significant differences. However, they do offer some interesting insights. In particular, one trend across the grades was the indication that male students were more motivated towards some of their CLIL subjects than female

learners. This is in keeping with the suggestion that CLIL may offer a more motivating context to male students who are interested in the subjects at hand.

Regarding French, the descriptive statistics for the language learning motivation in each language, as shown in Table 8.63, indicated that in 9th grade, female students reported higher motivation than male students in almost all categories and overall motivation, with the exception of *Language Anxiety*, where male students reported lower anxiety. Very different results were observed in 10th grade, where male students reported higher motivation than female students in almost all categories and overall motivation with the exception of *Interest in Foreign Languages*, where females reported higher interest. In 11th grade, male students again reported higher motivation than female students in almost all categories and overall motivation with the exception of *Interest in Foreign Languages*, where females reported higher interest, and *L2 Self Confidence*, where no difference was observed. In order to determine whether these differences were statistically significant, Mann-Whiney U tests were again conducted, comparing the results of the nine categories of the MFQ, as well as the overall motivation mean of the participants (Table 8.64). Results showed that there was a significant difference in the overall mean motivation between male students ($Mdn = 3.46, SD = .45$) and female students ($Mdn = 3.02, SD = .57$) only in 10th grade; $z = -2.44, p = .015$. This indicates that in 10th grade, male students appear to be more motivated towards learning French than female students, while in 9th and 11th grade there is no statistically significant difference. For the individual motivation categories, statistically significant differences were also found in two of the nine categories in 9th grade (*Instrumentality: Promotion*, *Instrumentality: Prevention*) and three of the nine categories in 10th grade (*Language Anxiety*, *Instrumentality: Promotion* and *Attitude towards Learning*). No statistically significant differences were found in 11th grade. Firstly, 9th grade male students ($Mdn = 2.83, SD = .99$) reported lower scores for *Instrumentality: Promotion* than female students ($Mdn = 3.50, SD = .85$); $z = -2.46, p = .014$. This suggests that the female students see French as more important in promoting their future success than male students. Secondly, 10th grade male students ($Mdn = 4.00, SD = .66$) reported lower anxiety than female students ($M = 3.30, SD = .98$); $z = -2.66, p = .008$. The male students ($Mdn = 3.25, SD = .66$) also reported a better attitude toward learning than female students ($M = 2.63, SD = .77$); $z = -2.99, p = .003$. Finally, as in the case of English, statistically significant differences were found at both levels for the category *Instrumentality: Prevention*, with 9th grade male students ($Mdn = 2.30, SD = .91$) reporting lower values than female students

($M = 3.00$, $SD = .71$); $z = -2.52$, $p = .012$, but 10th grade male students ($Mdn = 3.40$, $SD = .49$) reporting higher values than female students ($M = 2.90$, $SD = .71$); $z = -2.38$, $p = .017$. This indicates that in 9th grade, female students see not having French as preventing their future success to a greater extent than male students, while in 10th grade the reverse is true: male students see not having French as preventing their future success to a greater extent than female students. These results suggest some clear differences across levels in terms of gender. While in 9th grade, the female group was more instrumentality and intrinsically motivated than the male group, in 10th grade male students were more motivated than female students both overall, and particularly with regards to experiencing less anxiety towards learning French, believing that not having French would prevent their future success to a greater degree, and having a better attitude towards learning French. Meanwhile, no clear differences were found between male and female students in 11th grade. This is perhaps due to the fact that students at this level no longer took French CLIL classes, and so saw a reduction in their contact hours. This may have, to a certain extent, affected the students' motivation towards learning French. These results are presented visually in Figure 8.19.

Table 8.63*Descriptive Statistics for Language Learning Motivation by Gender in French*

9 th grade								
Male (<i>n</i> = 18)					Female (<i>n</i> = 23)			
Min	Max	<i>Mdn</i>	<i>SD</i>	Category	Min	Max	<i>Mdn</i>	<i>SD</i>
1.20	4.20	2.70	.98	Ideal L2 Self	1.80	5.00	2.80	.97
1.86	3.29	2.54	.42	The “Ought to” Self	1.14	3.57	2.57	.76
1.00	5.00	4.00	1.21	Language Anxiety	1.20	5.00	3.60	1.04
1.00	4.33	2.75	.97	Interest in FLs	1.50	5.00	3.17	.88
1.00	4.50	3.00	.86	L2 Self Confidence	2.25	4.50	3.50	.57
1.20	4.60	2.30	.91	Instrumentality: Prevention	1.80	4.80	3.00	.71
1.00	4.83	2.83	.99	Instrumentality: Promotion	1.33	4.83	3.50	.85
1.25	4.00	2.63	.79	Attitude towards Learning	1.25	4.50	2.38	.84
1.00	3.80	3.00	.70	Intended Learning Effort	1.40	5.00	3.20	1.03
1.77	3.48	2.89	.47	Mean Motivation	2.13	4.29	2.99	.56
10 th grade								
Male (<i>n</i> = 15)					Female (<i>n</i> = 26)			
Min	Max	<i>Mdn</i>	<i>SD</i>	Category	Min	Max	<i>Mdn</i>	<i>SD</i>
1.80	4.40	3.40	.77	Ideal L2 Self	1.00	4.40	2.90	.93
2.00	3.71	3.00	.52	The “Ought to” Self	1.57	3.57	2.57	.48
2.80	5.00	4.00	.66	Language Anxiety	1.20	5.00	3.30	.98
2.17	4.50	3.33	.62	Interest in FLs	1.67	4.17	3.42	.66
2.25	4.50	3.50	.65	L2 Self Confidence	1.00	5.00	3.25	.93
2.20	4.20	3.40	.49	Instrumentality: Prevention	1.20	4.00	2.90	.71
1.17	4.83	3.67	.85	Instrumentality: Promotion	1.67	4.67	3.50	.76
1.38	4.00	3.25	.66	Attitude towards Learning	1.00	3.75	2.63	.77
2.40	4.00	3.20	.61	Intended Learning Effort	1.20	4.40	2.90	.88
2.21	3.99	3.46	.45	Mean Motivation	1.61	4.04	3.02	.57
11 th grade								
Male (<i>n</i> = 15)					Female (<i>n</i> = 21)			
Min	Max	<i>Mdn</i>	<i>SD</i>	Category	Min	Max	<i>Mdn</i>	<i>SD</i>
1.60	4.20	3.20	.70	Ideal L2 Self	1.20	4.40	2.80	1.02
1.14	3.14	2.71	.54	The “Ought to” Self	1.14	3.86	2.57	.64
2.60	5.00	4.00	.72	Language Anxiety	1.80	5.00	3.60	1.10
2.25	4.33	3.33	.66	Interest in FLs	1.83	4.50	3.50	.72
2.25	4.50	3.25	.52	L2 Self Confidence	1.25	5.00	3.25	.88
1.40	4.00	3.00	.93	Instrumentality: Prevention	1.00	4.40	2.20	.92
1.67	4.83	3.50	.94	Instrumentality: Promotion	1.83	5.00	3.00	.87
1.63	4.00	3.13	.72	Attitude towards Learning	1.38	4.00	3.00	.56
1.60	4.00	3.20	.78	Intended Learning Effort	1.00	4.00	2.80	.80
2.15	3.91	3.16	.50	Mean Motivation	1.85	3.93	3.10	.54

Table 8.64*Differences in Language Learning Motivation by Gender in French*

9 th Grade						
Category	<i>Mdn</i>		<i>SD</i>		<i>z</i>	<i>p</i>
	Male	Female	Male	Female		
Ideal L2 Self	2.70	2.80	.98	.97	-1.19	.236
The “Ought to” Self	2.54	2.57	.42	.76	-.13	.895
Language Anxiety	4.00	3.60	1.21	1.04	-.98	.329
Interest in Foreign Languages	2.75	3.17	.97	.88	-1.78	.075
L2 Self Confidence	3.00	3.50	.86	.57	-1.56	.119
Instrumentality: Prevention	2.30	3.00	.91	.71	-2.52	.012
Instrumentality: Promotion	2.83	3.50	.99	.85	-2.46	.014
Attitude towards Learning	2.63	2.38	.79	.84	-1.01	.311
Intended Learning Effort	3.00	3.20	.70	1.03	-.41	.682
Mean French Motivation	2.89	2.99	.47	.56	-1.21	.227
10 th Grade						
Category	<i>Mdn</i>		<i>SD</i>		<i>z</i>	<i>p</i>
	Male	Female	Male	Female		
Ideal L2 Self	3.40	2.90	.77	.93	-1.14	.254
The “Ought to” Self	3.00	2.57	.52	.48	-1.52	.128
Language Anxiety	4.00	3.30	.66	.98	-2.66	.008
Interest in Foreign Languages	3.33	3.42	.62	.66	-0.48	.634
L2 Self Confidence	3.50	3.25	.65	.93	-1.07	.283
Instrumentality: Prevention	3.40	2.90	.49	.71	-2.38	.017
Instrumentality: Promotion	3.67	3.50	.85	.76	-0.91	.363
Attitude towards Learning	3.25	2.63	.66	.77	-2.99	.003
Intended Learning Effort	3.20	2.90	.61	.88	-0.87	.384
Mean French Motivation	3.46	3.02	.45	.57	-2.44	.015
11 th Grade						
Category	<i>Mdn</i>		<i>SD</i>		<i>z</i>	<i>p</i>
	Male	Female	Male	Female		
Ideal L2 Self	3.20	2.80	.70	1.02	-.60	.551
The “Ought to” Self	2.71	2.57	.54	.64	-.27	.784
Language Anxiety	4.00	3.60	.72	1.10	-.53	.595
Interest in Foreign Languages	3.33	3.50	.66	.72	-.32	.748
L2 Self Confidence	3.25	3.25	.52	.88	-.06	.948
Instrumentality: Prevention	3.00	2.20	.93	.92	-.64	.519
Instrumentality: Promotion	3.50	3.00	.94	.87	-.61	.541
Attitude towards Learning	3.13	3.00	.72	.56	-.50	.617
Intended Learning Effort	3.20	2.80	.78	.80	-1.05	.295
Mean French Motivation	3.16	3.10	.50	.54	-.53	.596

Figure 8.19

Differences in Language Learning Motivation by Gender in French



As in the English MFQ, analyses were run on the individual questions of each category in order to reveal any minor differences that may be found (Table 8.65). In 9th, 10th and 11th grade, 6, 13 and 3 out of the fifty-one individual questions, respectively, yielded statistically significant differences between the groups, with three questions common to 9th and 10th grade and one question common to 10th and 11th grade.

Table 8.65

Statistically Significant Differences between Groups in Individual Questions in French

9 th Grade						
Category	Mdn		SD		z	p
	Male	Fem	Male	Fem		
I need French to be a teacher.	2.00	3.00	0.98	1.21	-2.06	.039
I love hearing people speak French.	2.00	4.00	1.32	1.24	-2.78	.005
I don't want to fail with French because my professional future depends on it.	2.00	3.00	1.25	1.09	-2.00	.045
Studying French is important to me because I don't want to be seen as uncultured.	3.00	4.00	1.21	1.37	-2.16	.030
Learning French is important to me because I want to continue studying abroad.	2.00	3.00	1.33	0.97	-2.10	.035
I enjoy geography and history class.	4.00	2.00	1.50	1.21	-3.17	.002
10 th Grade						
Category	Mdn		SD		z	p
	Male	Fem	Male	Fem		
I can see myself living abroad and getting along with French people.	4.00	3.00	1.19	1.32	-2.38	.017
If a stranger asked me for directions, I would get nervous.	4.00	2.00	1.25	1.24	-3.13	.002
I feel silly in class when I speak French.	4.00	4.00	.77	1.17	-2.07	.038
I'm really curious about French structure and vocabulary.	3.00	2.00	.80	.80	-.57	<.001
Studying French is boring.	3.00	3.00	.98	1.13	-2.03	.042
I try to make the most of all situations to speak in French.	3.00	2.00	1.08	1.06	-1.22	.021
I don't want to fail with French because my professional future depends on it.	4.00	3.00	.85	1.12	-.03	.033
Not studying French would have a negative impact on my life.	4.00	3.00	.83	.95	-.56	.001
Learning French is important to me because I want to continue studying abroad.	4.00	2.50	1.12	1.20	-1.63	.042
Studying geography and history in French is really interesting.	3.00	2.00	1.26	.98	-1.83	.002
I enjoy French class.	3.00	2.00	.94	.97	-.32	.041
I enjoy geography and history class.	4.00	1.00	.99	1.17	-1.01	<.001
My experience in French class has always been positive.	4.00	3.00	.83	.99	-1.58	.015

Table 8.65 (continued)

11 th Grade						
Category	<i>Mdn</i>		<i>SD</i>		<i>z</i>	<i>p</i>
	Male	Fem	Male	Fem		
I love how French sounds.	3.00	4.00	1.10	1.21	-2.33	.020
Studying geography and history in French in secondary school was really interesting.	4.00	2.00	1.57	1.12	-2.26	.024
I will surely continue studying French when I finish school.	4.00	3.00	1.10	1.21	-2.77	.005

While these results should again be interpreted with caution, given that the overall category did not always show statistically significant differences, they offer some interesting insights. In 9th grade the results suggested that the female group had a higher score than the male group with regard to five specific statements: “I need French to be a teacher”, “I love hearing people speak French”, “I don’t want to fail with French because my professional future depends on it”, “Studying French is important to me because I don’t want to be seen as uncultured” and “Learning French is important to me because I want to continue studying abroad”. The male group had a higher score than the female group with regard to just one specific statements: “I enjoy geography and history class”. In 10th grade the results suggested that the male group had a higher score than the female group in all 13 specific statements which yielded statistically significant differences: “I can see myself living abroad and getting along with French people”, “If a stranger asked me for directions, I would get nervous”, “I feel silly in class when I speak French”, “I’m really curious about French structure and vocabulary”, “Studying French is boring”, “I try to make the most of all situations to speak in French”, “I don’t want to fail with French because my professional future depends on it”, “Not studying French would have a negative impact on my life”, “Learning French is important to me because I want to continue studying abroad”, “Studying geography and history in French is really interesting”, “I enjoy French class” and “I enjoy geography and history class”. As indicated by the results above, the analysis of the individual questions again suggest that the male group were more motivated and less anxious towards learning French than the female group with regards to these thirteen questions. Finally, in 11th grade the results suggested that the female group had a higher score than the male group with regard to just one specific statement: “I love how French sounds”. The male group had a higher score than the female group with regard to two specific statements: “Studying geography and history in French in secondary school was really interesting” and “I will surely

continue studying French when I finish school”. As compared to 10th grade group, which was made up of largely the same cohort of students, 11th grade male and female students notably had much fewer differences in their motivation towards French in the individual question, as well as in their overall motivation towards learning French. This, as suggested above, may be due to the reduction in French contact hours and the removal of CLIL classes in 11th grade. It is also again interesting to note that, as in the case of English, there appears to be a trend across the three grades indicating that male students were more motivated towards their French CLIL subject than female learners. This again offers interesting implications for the suggestion that CLIL provides a more motivating context to male students who are interested in the subjects at hand.

In summary, results revealed key differences between male and female students depending on the grade and language at hand (Table 8.66).

Table 8.66

Summary of Differences in Language Learning Motivation by Gender in English and French

	English			French		
	9 th	10 th	11 th	9 th	10 th	11 th
Ideal L2 Self						
The “Ought to” Self						
Language Anxiety					✓	
Interest in Foreign Languages						
L2 Self Confidence						
Instrumentality: Prevention	✓			✓		
Instrumentality: Promotion				✓	✓	
Attitude towards Learning					✓	
Intended Learning Effort						
Mean Motivation					✓	

In English, statistically significant differences were found only in the category *Instrumentality: Prevention* in 9th grade, with female students reporting that they saw not having English as preventing their future success to a greater degree than the male students. In French, however, clearer differences were found in 9th and 10th grade, where statistically significant differences were found in the categories *Instrumentality: Prevention* and *Instrumentality: Promotion* in 9th grade, and in the categories *Language Anxiety*, *Instrumentality: Prevention* and *Attitude towards Learning* and overall motivation in 10th grade. As in English, 9th grade female students saw not having French as preventing their future success to a greater degree than the male students, whereas in

10th grade it was male students who saw this to be more important. Female students in 9th grade also saw French as more important in promoting their future success than male students. Male students in 10th grade reported lower anxiety and a better attitude towards learning than female students.

Results of the analysis on the individual motivation questions also highlighted a potentially interesting issue: male students very often reported higher motivation towards CLIL subjects than female learners in both languages. In English, 9th grade male students reported finding learning biology and geology in English to be more interesting, 10th grade male students reported finding learning economics in English to be more interesting, and 11th grade students, reflecting on studying physical education in secondary school, reported finding studying this subject to be more interesting and reported that they had enjoyed it to a greater extent than female students. In French, 9th grade and 10th grade male students reported that they enjoyed studying geography and history to a greater extent than female students, and 11th grade students, reflecting on studying geography and history through French in secondary school, reported that they had found it more interesting than female students. These findings, though they should be interpreted with caution, offer interesting insights into the possibility that CLIL offers a more motivating context for male students who are interested in the subjects at hand.

As discussed in Section 8.3.3.1, to determine whether there were cross-sectional and longitudinal differences in the participants' language learning motivation in relation to gender and language, Mann-Whiney U tests were carried out to compare students in 9th grade to students in 10th grade and Wilcoxon signed-rank tests were carried out to compare students in 10th grade and 11th grade. The results are analysed here with a focus on comparing the results of male versus female students in English, and male versus female students in French.

Firstly, regarding the cross-sectional analysis, results revealed that in the case of English language learning motivation there were no statistically significant differences between 9th grade and 10th grade students for male students (Table 8.67), while for female students a statistically significant difference was found between 9th grade and 10th grade students only in the category *The "Ought to" Self*. In this case, 10th grade female students reported higher motivation due to external sources than their 9th grade peers.

Table 8.67*Cross-Sectional Differences in Language Learning Motivation in English by Gender*

9 th → 10 th grade: Male						
	<i>Mdn</i>		<i>SD</i>		<i>z</i>	<i>p</i>
	9 th	10 th	9 th	10 th		
Ideal L2 Self	3.80	4.20	.888	.506	-1.07	.283
The “Ought to” Self	2.57	3.00	.566	.551	-1.35	.176
Language Anxiety	4.40	4.20	.905	.891	-0.54	.590
Interest in Foreign Languages	3.83	4.00	.885	.580	-0.24	.811
L2 Self Confidence	3.88	3.75	.778	.837	-0.66	.510
Instrumentality: Prevention	3.15	3.60	.084	.612	-1.23	.219
Instrumentality: Promotion	3.96	4.33	.886	.425	-0.72	.474
Attitude towards Learning	3.60	3.60	.837	.473	-0.71	.476
Intended Learning Effort	3.80	3.80	.575	.504	-0.34	.734
Mean English Motivation	3.47	3.79	.512	.402	-0.91	.363
9 th → 10 th grade: Female						
	<i>Mdn</i>		<i>SD</i>		<i>z</i>	<i>p</i>
	9 th	10 th	9 th	10 th		
Ideal L2 Self	4.20	3.80	.761	.699	-0.62	.537
The “Ought to” Self	2.57	3.00	.602	.409	-2.43	.015
Language Anxiety	3.80	3.80	.914	.913	-0.10	.920
Interest in Foreign Languages	3.83	3.67	.617	.607	-1.32	.186
L2 Self Confidence	3.75	3.63	.638	.800	-1.07	.285
Instrumentality: Prevention	4.20	4.00	.486	.662	-0.04	.968
Instrumentality: Promotion	4.20	4.00	.605	.512	-1.48	.138
Attitude towards Learning	3.36	3.55	.818	.547	-0.52	.602
Intended Learning Effort	4.00	3.80	.625	.569	-1.84	.066
Mean English Motivation	3.72	3.73	.445	.416	-0.58	.561

On the other hand, in the case of French language learning motivation (Table 8.68), statistically significant differences between 9th grade and 10th grade students for male students were found in the participants’ mean French language learning motivation as well as a number of categories: *The “Ought to” Self*, *Interest in Foreign Languages*, *Instrumentality: Prevention* and *Instrumentality: Promotion*. In all cases, means were higher for 10th grade students than for 9th grade students. Meanwhile, for female participants no statistically significant differences were found between 9th grade and 10th grade students. These results highlight clear differences between male and female students’ language learning motivation depending on the language at hand. While few differences arise in English, there are very clear differences in French between 9th and 10th grade male students, but not between 9th and 10th grade female students. As discussed above, 10th grade male students report higher motivation due to external sources, greater

interest in learning French, see French as more important in promoting their future success, and see not having French as more preventative to their future success. However, the same is not seen for female participants, who are equally motivated towards French in 9th and 10th grade. There thus appears to be a greater difference from one grade to the next for male students than for female students in French, but not in English.

Table 8.68*Cross-Sectional Differences in Language Learning Motivation in French by Gender*

9 th → 10 th grade: Male						
	<i>Mdn</i>		<i>SD</i>		<i>z</i>	<i>p</i>
	9 th	10 th	9 th	10 th		
Ideal L2 Self	2.70	3.40	.979	.767	-1.96	.050
The “Ought to” Self	2.54	3.00	.416	.519	-1.98	.048
Language Anxiety	4.00	4.00	1.21	.656	-0.51	.611
Interest in Foreign Languages	2.75	3.33	.967	.620	-2.14	.032
L2 Self Confidence	3.00	3.50	.857	.651	-1.53	.127
Instrumentality: Prevention	2.30	3.40	.915	.493	-3.08	.002
Instrumentality: Promotion	2.83	3.67	.992	.851	-2.65	.008
Attitude towards Learning	2.63	3.25	.788	.662	-1.98	.048
Intended Learning Effort	3.00	3.20	.702	.609	-0.24	.813
Mean French Motivation	2.89	3.46	.465	.447	-3.29	.001
9 th → 10 th grade: Female						
	<i>Mdn</i>		<i>SD</i>		<i>z</i>	<i>p</i>
	9 th	10 th	9 th	10 th		
Ideal L2 Self	2.80	2.90	.969	.935	-0.08	.936
The “Ought to” Self	2.57	2.57	.764	.479	-0.62	.534
Language Anxiety	3.60	3.30	1.04	.984	-0.60	.547
Interest in Foreign Languages	3.17	3.42	.878	.657	-0.22	.825
L2 Self Confidence	3.50	3.25	.569	.932	-0.94	.349
Instrumentality: Prevention	3.00	2.90	.707	.709	-0.95	.344
Instrumentality: Promotion	3.50	3.50	.852	.759	-0.44	.658
Attitude towards Learning	2.38	2.63	.838	.768	-0.04	.968
Intended Learning Effort	3.20	2.90	1.03	.875	-0.91	.361
Mean French Motivation	2.99	3.02	.558	.572	-0.18	.857

Secondly, regarding the longitudinal analysis, results revealed that in the case of English language learning motivation there was a statistically significant difference from 10th to 11th grade in the male participants’ overall language learning motivation and in the category *The “Ought to” Self*, while for female students a statistically significant difference was found from 10th to 11th grade only in the category *Attitude towards Learning* (Table 8.69).

Table 8.69*Longitudinal Differences in Language Learning Motivation in English by Gender*

10 th → 11 th grade: Male						
	<i>Mdn</i>		<i>SD</i>		<i>Z</i>	<i>p</i>
	10 th	11 th	10 th	11 th		
Ideal L2 Self	4.20	4.20	.506	.715	-.444	.657
The “Ought to” Self	3.00	3.00	.551	.518	-2.14	.032
Language Anxiety	4.20	4.20	.891	.888	-.079	.937
Interest in Foreign Languages	4.00	3.83	.580	.591	-1.77	.076
L2 Self Confidence	3.75	3.75	.837	.680	-.694	.487
Instrumentality: Prevention	3.60	3.80	.612	.648	-.396	.692
Instrumentality: Promotion	4.33	4.33	.425	.651	-.599	.549
Attitude towards Learning	3.60	3.42	.473	.689	-1.85	.064
Intended Learning Effort	3.80	3.80	.504	.723	-1.91	.056
Mean English Motivation	3.79	3.80	.402	.435	-2.04	.041
10 th → 11 th grade: Female						
	<i>Mdn</i>		<i>SD</i>		<i>Z</i>	<i>p</i>
	10 th	11 th	10 th	11 th		
Ideal L2 Self	3.80	4.40	.699	.666	-1.00	.316
The “Ought to” Self	3.00	3.00	.409	.684	-.029	.977
Language Anxiety	3.80	4.00	.913	.854	-.380	.704
Interest in Foreign Languages	3.67	3.83	.607	.667	-.168	.867
L2 Self Confidence	3.63	3.75	.800	.917	-.493	.622
Instrumentality: Prevention	4.00	4.00	.662	.611	-.625	.532
Instrumentality: Promotion	4.00	4.00	.512	.493	-.190	.849
Attitude towards Learning	3.55	3.10	.547	.628	-2.74	.006
Intended Learning Effort	3.80	4.00	.569	.527	-.211	.833
Mean English Motivation	3.73	3.78	.416	.389	-.327	.744

Similarly, in the case of French language learning motivation (Table 8.70), there was a statistically significant difference from 10th to 11th grade in the male participants' overall language learning motivation and in the category *The “Ought to” Self*. In all cases, means were lower in 11th grade than in 10th grade, suggesting a decrease in language learning motivation. On the other hand, for female students, no statistically significant differences were found from 10th to 11th grade. As in the case of the cross-sectional analysis, these results suggest that there is a greater difference from one grade to the next for male students than for female students, however, while these differences are seen in fewer categories, they are seen for both languages. In particular, 11th grade male students report lower motivation overall and lower motivation due to external sources than 10th grade male students in both languages, while female learners report a less positive attitude towards learning English in 11th grade than in 10th grade.

Table 8.70*Longitudinal Differences in Language Learning Motivation in French by Gender*

10 th → 11 th grade: Male						
	<i>Mdn</i>		<i>SD</i>		<i>Z</i>	<i>p</i>
	10 th	11 th	10 th	11 th		
Ideal L2 Self	3.40	3.20	.767	.705	-1.59	.111
The “Ought to” Self	3.00	2.71	.519	.544	-2.09	.036
Language Anxiety	4.00	4.00	.656	.718	-1.20	.227
Interest in Foreign Languages	3.33	3.33	.620	.661	-1.06	.288
L2 Self Confidence	3.50	3.25	.651	.516	-.632	.528
Instrumentality: Prevention	3.40	3.00	.493	.934	-1.96	.050
Instrumentality: Promotion	3.67	3.50	.851	.942	-1.61	.107
Attitude towards Learning	3.25	3.13	.662	.725	-1.72	.085
Intended Learning Effort	3.20	3.20	.609	.778	-1.01	.311
Mean French Motivation	3.46	3.16	.447	.502	-2.04	.041
10 th → 11 th grade: Female						
	<i>Mdn</i>		<i>SD</i>		<i>Z</i>	<i>p</i>
	10 th	11 th	10 th	11 th		
Ideal L2 Self	2.90	2.80	.935	1.01	-.052	.959
The “Ought to” Self	2.57	2.57	.479	.644	-.591	.555
Language Anxiety	3.30	3.60	.984	1.10	-1.27	.204
Interest in Foreign Languages	3.42	3.50	.657	.717	-.119	.905
L2 Self Confidence	3.25	3.25	.932	.879	-1.06	.289
Instrumentality: Prevention	2.90	2.20	.709	.924	-1.67	.095
Instrumentality: Promotion	3.50	3.00	.759	.869	-1.83	.066
Attitude towards Learning	2.63	3.00	.768	.563	-1.19	.231
Intended Learning Effort	2.90	2.80	.875	.798	-1.62	.104
Mean French Motivation	3.02	3.10	.572	.544	-.283	.777

8.4. Content and Language Integrated Learning

Research question 4 aimed to determine the effect of CLIL instruction on the language learning of secondary school students of English and French, in particular concerning quantitative differences in the language level, LA and motivation of students taking different CLIL classes (RQ4.1), the qualitative differences in the words produced in English by students taking different CLIL classes at each testing period (RQ4.2), and the quantitative (RQ4.3.) and qualitative (RQ4.4) differences in the words produced by students in English and French when the prompt is related to a CLIL class taken in either English or French at each testing period.

8.4.1. Quantitative Differences in CLIL: Language Level, Lexical Availability and Motivation

Research question 4.1 asked if there were quantitative differences in the language level, LA and motivation of students taking different CLIL classes. To this effect, students in 10th and 11th grade were divided into groups depending on the English CLIL class (a science subject or economics) they had chosen. In 10th grade this included those taking a science subject in English (Physics and Chemistry, or Biology) ($n = 26$) and those taking economics in English ($n = 15$), and in 11th grade this included those taking the same science subject in Spanish (Physics and Chemistry, or Biology) ($n = 21$) and those taking economics in English or Spanish ($n = 19$; 10 through English and 9 through Spanish). In the longitudinal analysis, only students who studied the same language in both grades were included (Physics and Chemistry, or Biology: $n = 18$, Economics: $n = 10$). Given that students in 9th grade all studied the same CLIL subjects, they were not included in this analysis. The analysis was carried out first for language level and LA and then then for language learning motivation.

Firstly, regarding language level and LA, the descriptive statistics for the participants' English language level and LA in English, as shown in Table 8.71, indicated that science students had a higher English language level than economics students and produced a higher number of tokens in the overall English LAT and almost all prompts, with just one exception: *Economy & Money* in 10th grade. In order to determine if these differences were statistically significant, the normality of the results of the English C-test and LAT were first assessed for each subgroup of students taking different CLIL classes in 10th and 11th grade (Table 8.72). As shown, data were normally distributed in all cases. Thus, independent sample t-tests were carried out to compare these groups of students in terms of their performance on the English C-test to determine if there were differences in their English language level, and in terms of their performance on the English LAT to determine if there were differences in their LA (Table 8.73). The results are presented visually in Figure 8.20.

Table 8.71*Descriptive Statistics for English Language Level and LA by CLIL Group*

10 th Grade								
Science (<i>n</i> = 26)					Economics (<i>n</i> = 15)			
Min	Max	<i>M</i>	<i>SD</i>	Prompt	Min	Max	<i>M</i>	<i>SD</i>
22	60	38.00	10.43	C-test English	5	47	27.20	11.37
13	28	20.08	4.78	Animals	10	19	15.00	2.90
13	35	22.92	5.43	Food & Drink	12	27	19.33	4.34
11	27	18.38	4.19	Sport & PA	9	19	13.53	3.50
11	31	20.42	4.97	Env.t & Climate	2	26	15.47	6.13
4	26	13.85	4.95	Economy & Money	4	22	14.13	4.67
12.80	29.20	19.13	3.72	Mean LAT	9.20	21.40	15.49	3.50
11 th Grade								
Science (<i>n</i> = 21)					Economics (<i>n</i> = 19)			
Min	Max	<i>M</i>	<i>SD</i>	Prompt	Min	Max	<i>M</i>	<i>SD</i>
24	63	44.10	10.71	C-test English	17	52	30.32	10.36
14	30	23.43	4.60	Animals	7	23	16.16	4.07
11	36	25.52	6.58	Food & Drink	8	30	19.68	6.79
10	27	19.29	4.54	Sport & PA	9	21	15.21	3.84
8	32	21.24	6.74	Env. & Climate	4	29	15.26	6.97
2	24	14.10	5.73	Economy & Money	7	20	13.53	3.99
10.80	28.40	20.71	4.42	Mean LAT	9.60	23.80	15.97	4.02

Table 8.72*Normality Tests for English C-test and LAT by CLIL group*

10 th grade						
Science				Economics		
<i>W</i>	<i>df</i>	<i>p</i>		<i>W</i>	<i>df</i>	<i>p</i>
.943	26	.161	C-test English	.978	15	.951
.942	26	.151	Animals	.904	15	.108
.974	26	.733	Food & Drink	.938	15	.359
.933	26	.091	Sport & Physical Activities	.903	15	.106
.974	26	.736	Environment & Climate	.968	15	.825
.971	26	.641	Economy & Money	.978	15	.951
.962	26	.426	Mean LAT	.969	15	.849
11 th grade						
Science				Economics		
<i>W</i>	<i>df</i>	<i>p</i>		<i>W</i>	<i>df</i>	<i>p</i>
.975	21	.840	C-test English	.939	19	.254
.937	21	.188	Animals	.946	19	.338
.957	21	.466	Food & Drink	.952	19	.429
.943	21	.253	Sport & Physical Activities	.920	19	.111
.967	21	.662	Environment & Climate	.973	19	.835
.975	21	.841	Economy & Money	.949	19	.381
.957	21	.463	Mean LAT	.957	19	.509

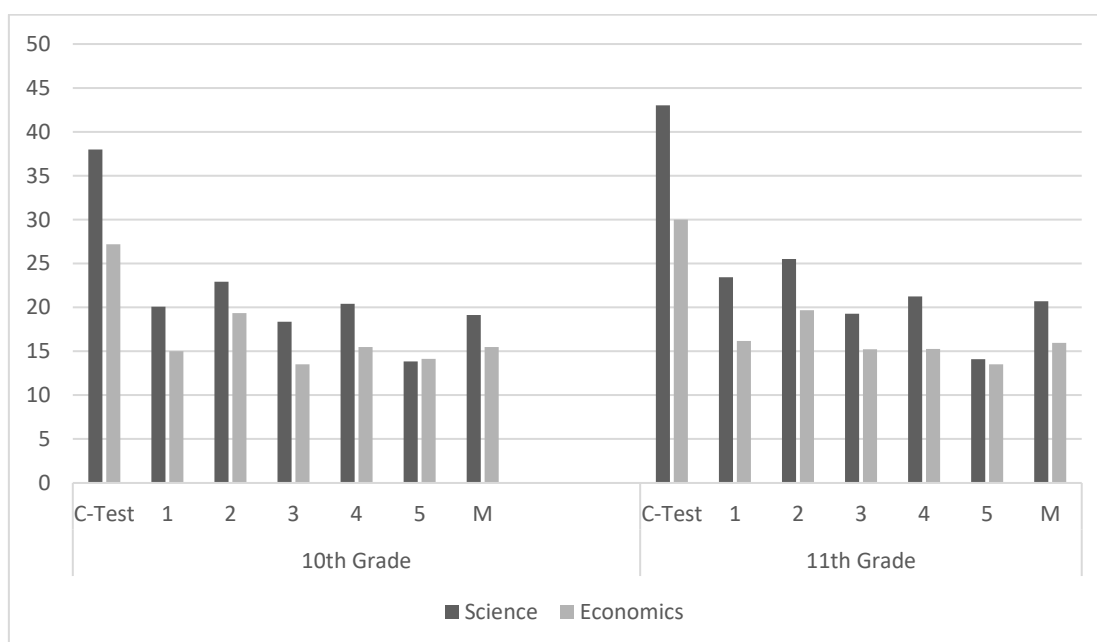
Table 8.73

Comparison of the Groups “Science” and “Economics”: Language Level and LAT

10 th grade							
Category	M		SD		<i>t</i>	<i>df</i>	<i>p</i>
	Sci	Eco	Sci	Eco			
C-Test English	38.00	27.20	10.42	11.36	3.09	39	.004
Animals	20.08	15.00	4.78	2.90	4.22	38.84	<.001
Food & Drink	22.92	19.33	5.42	4.33	2.18	39	.035
Sport & Physical Activities	18.38	13.53	4.18	3.50	3.78	39	.001
Environment & Climate	20.42	15.47	4.96	6.12	2.82	39	.007
Economy & Money	13.85	14.13	4.94	4.67	-.183	39	.856
Mean LAT	19.13	15.49	3.71	3.50	3.08	39	.004
11 th grade							
Category	M		SD		<i>t / z</i>	<i>df</i>	<i>p</i>
	Sci	Eco	Sci	Eco			
C-Test English	44.10	30.32	10.70	10.36	4.12	38	<.001
Animals	23.43	16.16	4.60	4.07	5.26	38	<.001
Food & Drink	25.52	19.68	6.58	6.79	2.76	38	.009
Sport & Physical Activities	19.29	15.21	4.54	3.83	3.04	38	.004
Environment & Climate	21.24	15.26	6.73	6.97	2.75	38	.009
Economy & Money	14.10	13.53	5.72	3.99	.361	38	.720
Mean LAT	20.71	15.96	4.42	4.01	3.54	38	.001

Figure 8.20

English Language Level and LAT by CLIL group



Note. 1 = Animals, 2 = Food & Drink, 3 = Sport and Physical Activities, 4 = Environment & Climate, 5 = Economy & Money, M = Mean LAT.

It should first be noted that science students in both 10th grade and 11th grade were found to have a statistically significant higher level of English than the economics students in terms of their performance on the C-test. Secondly, results showed that the science students produced a statistically significant higher number of words in the overall LAT, as well as in four out of the five prompts, again in both grades. However, it is very interesting to note that, despite the economics students having a lower level of English and producing statistically fewer words than the science students in all other categories, there was no statistically significant difference between the two groups in the prompt *Economy & Money* in 10th or 11th grade. This finding suggests that, to some extent, by studying economics through English, the students have improved their content-related vocabulary in this area to the extent that they are not outperformed by the science students, who appear to have a higher level of English and also outperform them in all other categories and overall.

In order to determine if there were longitudinal differences between the science ($n = 18$) and economics students ($n = 10$) from 10th to 11th grade, two-way mixed ANOVAs were carried out to compare the effect of time on CLIL group, comparing each groups' performance on the English C-test and their performance on the English LAT in each grade (Table 8.74). As noted above, only students who studied the same subject in each grade were included, excluding the four longitudinal participants who changed from studying science in 10th grade to studying economics in 11th grade or vice versa.

Regarding language level, results indicated that there was a significant main effect of CLIL group on the English C-test at each time ($F(1, 26) = 14.83, p = .001, \eta_p^2 = .363$), with the science group ($M = 41.55$) performing better than the economics group ($M = 25.50$) overall. There was also a significant main effect of time ($F(1, 26) = 4.51, p = .043, \eta_p^2 = .148$), with participants improving from 10th grade ($M = 32.01$) to 11th grade ($M = 35.04$). However, there was no significant interaction between time and CLIL group ($F(1, 26) = .197, p = .661, \eta_p^2 = .008$); the science group increased their score from 10th grade ($M = 39.72$) to 11th grade ($M = 43.38$), and the economics group increased their mean from 10th grade ($M = 24.30$) to 11th grade ($M = 26.70$).

Regarding the English LAT, results indicated that there was a significant main effect of CLIL group on the overall English LAT at each time ($F(1, 26) = 10.06, p = .004, \eta_p^2 = .279$), with the science group ($M = 20.23$) producing more tokens than the economics group ($M = 15.13$) overall. There was, however, no significant main effect of time ($F(1, 26) = .770, p = .388, \eta_p^2 = 0.29$), with means in 10th grade ($M = 17.44$)

increasing only slightly in 11th grade ($M = 17.91$). There was also no significant interaction between time and CLIL group ($F(1, 26) = .588, p = .450, \eta_p^2 = .022$); the science group increased their mean from 10th grade ($M = 19.78$) to 11th grade ($M = 20.67$), and the economics group increased their mean from 10th grade ($M = 15.10$) to 11th grade ($M = 15.16$). Regarding the individual prompts of the LAT, there was a significant main effect of CLIL group in the first four prompts at each time, with science students producing more tokens than economics students in all cases (*Animals*: $F(1, 26) = 20.48, p = <.001, \eta_p^2 = .441$; *Food & Drink*: $F(1, 26) = 10.56, p = .003, \eta_p^2 = .289$; *Sport & Physical Activity*: $F(1, 26) = 13.90, p = .001, \eta_p^2 = .348$; *Environment & Climate*: $F(1, 26) = 7.00, p = .014, \eta_p^2 = .212$). However, it is very notable that with regard to the prompt *Economy & Money*, no significant main effect of CLIL group was observed ($F(1, 26) = .000, p = .986, \eta_p^2 = .000$). This indicates that while science students performed better from 10th to 11th grade than economics students in all other cases, in the case of this content-relevant prompt, both groups performed similarly. The results also indicated that, as in the case of the overall LAT, there was no significant main effect of time on any of the individual prompts (*Animals*: $F(1, 26) = .283, p = .104, \eta_p^2 = .098$; *Food & Drink*: $F(1, 26) = .322, p = .575, \eta_p^2 = 0.12$; *Sport & Physical Activity*: $F(1, 26) = .076, p = .785, \eta_p^2 = .003$; *Environment & Climate*: $F(1, 26) = .121, p = .731, \eta_p^2 = .005$; *Economy & Money*: $F(1, 26) = .158, p = .694, \eta_p^2 = .006$) and that there was no significant interaction between time and CLIL group for any of the five prompts of the English LAT (*Animals*: $F(1, 26) = 1.66, p = .208, \eta_p^2 = 0.60$); *Food & Drink*: $F(1, 26) = .703, p = .409, \eta_p^2 = .026$; *Sport & Physical Activity*: $F(1, 26) = .076, p = .785, \eta_p^2 = .003$; *Environment & Climate*: $F(1, 26) = 1.69, p = .205, \eta_p^2 = .061$; *Economy & Money*: $F(1, 26) = 1.17, p = .289, \eta_p^2 = .043$). These results show that there was significant main effect of CLIL group on the English C-test, the overall English LAT, and four out of the five individual prompts of the English LAT at each time. In all cases, science students performed better than economics students. However, with regard to the content-relevant prompt, *Economy & Money*, both groups performed similarly. A significant main effect of time was found for the English C-test, but not the overall English LAT, or any of the five individual prompts, suggesting that while both groups' language level improved from 10th to 11th grade, there was no clear difference in the number of tokens produced from one grade to the next. There were also no significant interactions between time and CLIL group for the English C-test, the overall English LAT, or any of the five individual prompts, suggesting that

performance at each grade was not influenced by CLIL group. These results are represented visually in Figure 8.21.

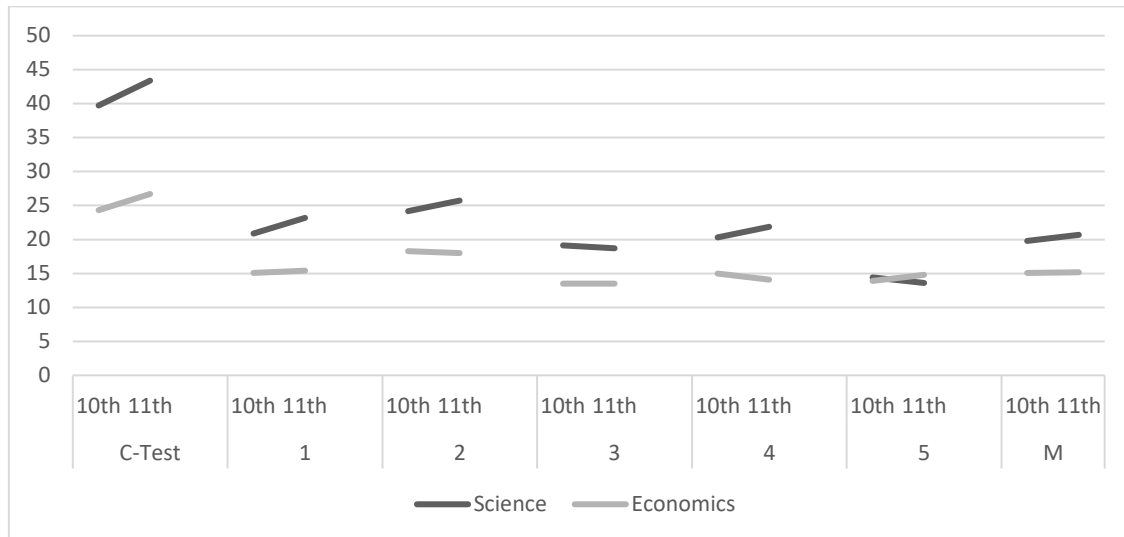
Table 8.74

Longitudinal Differences in the Groups “Science” and “Economics”: Language Level and LAT

CLIL group							
	M				F	p	η_p^2
	Science		Economics				
C-Test	41.556		25.50		14.83	.001	.363
1	22.02		15.25		20.48	<.001	.441
2	24.94		18.15		10.56	.003	.289
3	18.91		13.50		13.90	.001	.348
4	21.11		14.55		7.009	.014	.212
5	14.16		14.20		.000	.986	.000
Mean	20.23		15.13		10.06	.004	.279
Time							
	M				F	p	η_p^2
	10 th Grade		11 th Grade				
C-Test	32.01		35.04		4.51	.043	.148
1	17.99		19.28		2.83	.104	.098
2	21.23		21.86		.322	.575	.012
3	16.30		16.11		.076	.785	.003
4	17.66		17.99		.121	.731	.005
5	14.02		14.34		.158	.694	.006
Mean	17.44		17.91		.770	.388	.029
CLIL Group*Time							
	M				F	p	η_p^2
	10 th Grade		11 th Grade				
	Science	Economics	Science	Economics			
C-Test	39.72	24.30	43.38	26.70	.197	.661	.008
1	20.88	15.10	23.16	15.40	1.66	.208	.060
2	24.16	18.30	25.72	18.00	.703	.409	.026
3	19.11	13.50	18.72	13.50	.076	.785	.003
4	20.33	15.00	21.88	14.10	1.69	.205	.061
5	14.44	13.88	13.60	14.800	1.17	.289	.043
Mean	19.78	15.10	20.67	15.16	.588	.450	.022

Figure 8.21

Longitudinal Differences between the Groups “Science” and “Economics”: Language Level and LAT



Note. 1 = *Animals*, 2 = *Food & Drink*, 3 = *Sport and Physical Activities*, 4 = *Environment & Climate*, 5 = *Economy & Money*, M = Mean LAT.

Secondly, regarding motivation, the descriptive statistics for the participants' English language learning motivation, as shown in Table 8.75, indicated that in 10th grade science students reported higher motivation than economics students overall and in almost all categories with the exception of *The “Ought to” Self* and *L2 Self Confidence*. In 11th grade, economics students reported higher motivation than science students overall and in the category *The “Ought to” Self*, however, the majority of the other categories had either higher medians for science students (*Language Anxiety*, *L2 Self Confidence*, *Instrumentality: Promotion*, *Attitude towards Learning*) or similar medians for both groups (*Ideal L2 Self*, *Interest in Foreign Languages*, *Instrumentality: Prevention*). Given that the data at hand were ordinal, non-parametric Mann-Whitney U tests were carried out to determine if these differences between the science and economics students in their English language learning motivation were statistically significant. As shown in Table 8.76 below, no statistically significant differences were found between in the two groups in their overall English language learning motivation in either 10th or 11th grade. In addition, no differences were found in any of the individual categories in 10th grade, suggesting that the linguistic advantage of the science students discussed above may not be attributable to motivation. Few differences were found in 11th grade, with statistically significant differences found only in the categories *The “Ought to” Self* and *L2 Self*

Confidence. The results indicated that the economics students were more motivated by external sources but had lower self-confidence in English than the science students. However, it appears that given that there were few differences between the two groups, the linguistic advantage of the science students in 11th grade may again not be attributed to higher motivation. These findings are represented visually in Figure 8.22.

Table 8.75*Descriptive Statistics for English Language Learning Motivation by CLIL Group*

10 th Grade									
Science (<i>n</i> = 26)					Economics (<i>n</i> = 15)				
Min	Max	<i>Mdn</i>	<i>SD</i>		Min	Max	<i>Mdn</i>	<i>SD</i>	
3.20	5.00	4.20	.54	Ideal L2 Self	2.60	5.00	3.80	.74	
2.14	3.86	2.86	.48	The “Ought to” Self	2.71	3.86	3.43	.38	
1.60	5.00	3.90	.93	Language Anxiety	2.00	5.00	3.80	.89	
3.33	4.67	3.83	.46	Interest in Foreign Languages	2.17	4.83	3.50	.74	
2.00	5.00	3.63	.77	L2 Self Confidence	2.00	4.33	3.75	.80	
2.40	5.00	4.00	.71	Instrumentality: Prevention	2.40	5.00	3.80	.62	
3.00	5.00	4.17	.46	Instrumentality: Promotion	3.17	5.00	4.00	.53	
2.50	4.40	3.60	.44	Attitude towards Learning	2.11	4.30	3.30	.64	
2.80	4.80	3.80	.57	Intended Learning Effort	2.40	4.40	3.60	.49	
3.17	4.42	3.82	.39	Mean English Motivation	2.9	4.3	3.74	.44	
11 th Grade									
Science (<i>n</i> = 21)					Economics (<i>n</i> = 19)				
Min	Max	<i>Mdn</i>	<i>SD</i>		Min	Max	<i>Mdn</i>	<i>SD</i>	
2.60	5.00	4.20	.70	Ideal L2 Self	2.60	5.00	4.20	.69	
1.71	3.43	2.83	.55	The “Ought to” Self	2.00	4.43	3.14	.58	
2.00	5.00	4.20	.92	Language Anxiety	2.80	5.00	4.00	.81	
2.50	4.83	3.83	.61	Interest in Foreign Languages	2.50	5.00	3.83	.65	
2.75	5.00	4.00	.66	L2 Self Confidence	1.75	5.00	3.50	.87	
2.80	5.00	4.00	.72	Instrumentality: Prevention	2.60	5.00	4.00	.55	
3.17	5.00	4.17	.56	Instrumentality: Promotion	2.67	5.00	4.00	.59	
2.00	4.30	3.40	.60	Attitude towards Learning	1.75	4.60	3.20	.72	
1.80	4.80	4.00	.73	Intended Learning Effort	2.20	4.40	3.80	.55	
3.02	4.34	3.69	.41	Mean English Motivation	2.77	4.59	3.85	.42	

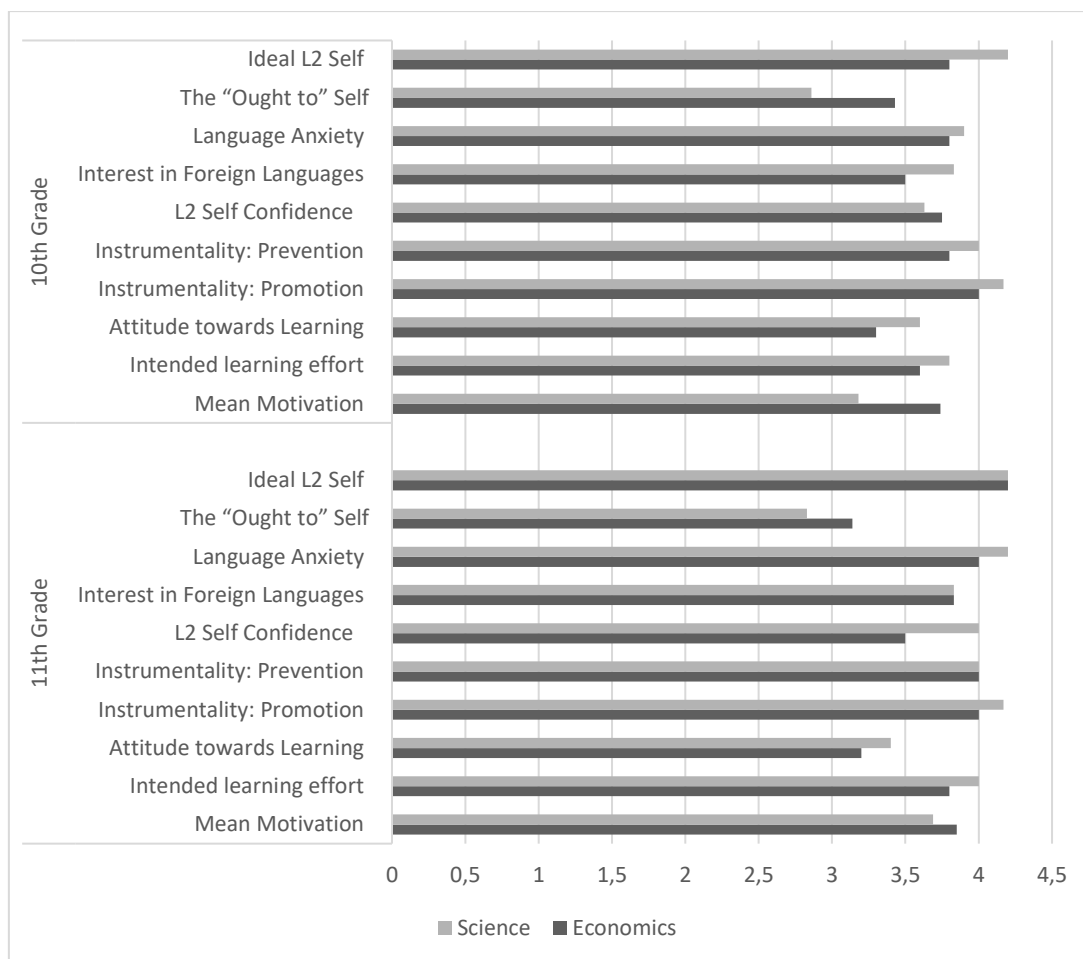
Table 8.76*Comparison of the Groups “Science” and “Economics”: English Language Motivation*

10 th Grade						
Category	<i>M</i>		<i>SD</i>		<i>z</i>	<i>p</i>
	Sci	Eco	Sci	Eco		
Ideal L2 Self	4.20	3.80	.54	.74	-1.13	.257
The “Ought to” Self	2.86	3.43	.48	.38	-1.79	.074
Language Anxiety	3.90	3.80	.93	.89	-.24	.807
Interest in Foreign Languages	3.83	3.50	.46	.74	-1.73	.084
L2 Self Confidence	3.63	3.75	.77	.80	-1.46	.145
Instrumentality: Prevention	4.00	3.80	.71	.62	-.10	.924
Instrumentality: Promotion	4.17	4.00	.46	.53	-.05	.956
Attitude towards Learning	3.60	3.30	.44	.64	-1.19	.232
Intended Learning Effort	3.80	3.60	.57	.49	-.57	.567
Mean English Motivation	3.81	3.74	.39	.44	-1.16	.244
11 th Grade						
Category	<i>M</i>		<i>SD</i>		<i>z</i>	<i>p</i>
	Sci	Eco	Sci	Eco		
Ideal L2 Self	4.20	4.20	.70	.69	-.12	.903
The “Ought to” Self	2.83	3.14	.55	.58	-2.08	.037
Language Anxiety	4.20	4.00	.92	.81	-.04	.967
Interest in Foreign Languages	3.83	3.83	.61	.65	-.37	.713
L2 Self Confidence	4.00	3.50	.66	.87	-2.01	.044
Instrumentality: Prevention	4.00	4.00	.72	.55	-.15	.881
Instrumentality: Promotion	4.17	4.00	.56	.59	-.30	.764
Attitude towards Learning	3.40	3.20	.60	.72	-1.00	.316
Intended Learning Effort	4.00	3.80	.73	.55	-.29	.774
Mean English Motivation	3.69	3.85	.41	.42	-.27	.787

Note. Science ($n = 26$) and Economics ($n = 15$) in 10th grade; Science ($n = 21$) and Economics ($n = 19$) in 11th grade.

Figure 8.22

English Language Motivation by CLIL Group



In order to determine if there were longitudinal differences between the science ($n = 18$) and economics students ($n = 10$) from 10th to 11th grade, Kruskal-Wallis H tests were carried out to compare the groups' English language learning motivation at each point. The results are presented in Table 8.77. As shown, no statistically significant differences were found between the groups in either 10th or 11th grade, again suggesting that there are no longitudinal differences in English language learning motivation which are dependent on CLIL group. These results are presented visually in Figure 8.23.

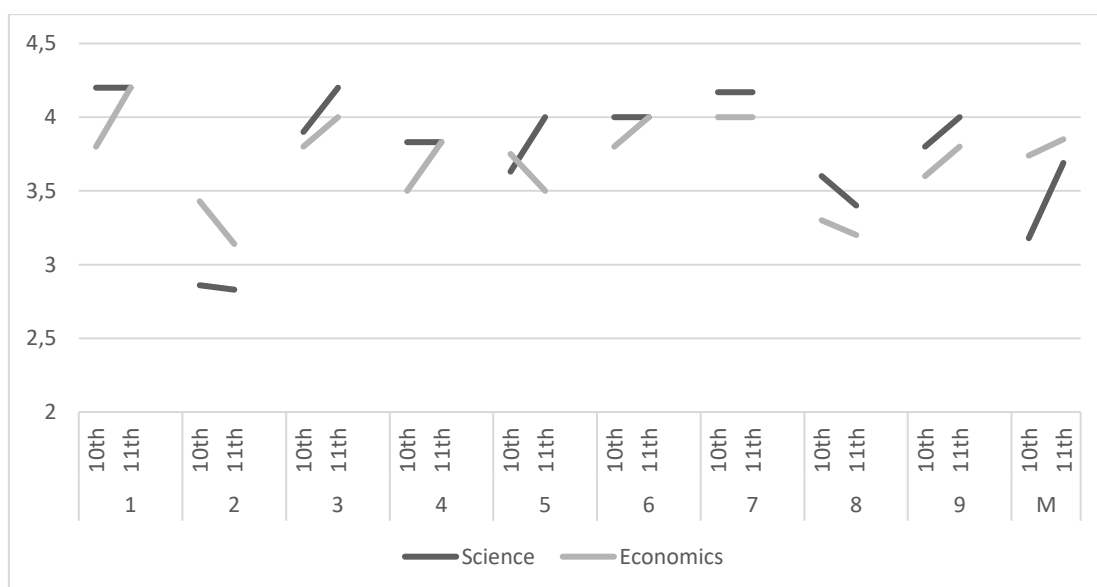
Table 8.77

Longitudinal Differences between the Groups “Science” and “Economics”: English Language Motivation

	10 th Grade			11 th Grade		
	<i>H</i>	<i>Df</i>	<i>p</i>	<i>H</i>	<i>Df</i>	<i>p</i>
Ideal L2 Self	.283	1	.595	.084	1	.773
The “Ought to” Self	2.239	1	.135	1.817	1	.178
Language Anxiety	.336	1	.562	1.249	1	.264
Interest in Foreign Languages	.363	1	.547	.114	1	.736
L2 Self Confidence	1.133	1	.287	1.464	1	.226
Instrumentality: Prevention	.559	1	.455	.130	1	.718
Instrumentality: Promotion	.071	1	.790	.189	1	.664
Attitude towards Learning	.070	1	.791	1.557	1	.212
Intended Learning Effort	.233	1	.629	.015	1	.903
Mean English Motivation	.278	1	.598	.047	1	.829

Figure 8.23

Longitudinal Differences between the Groups “Science” and “Economics”: English Language Motivation



Note. 1 = Ideal L2 Self, 2 = The “Ought to” Self, 3 = Language Anxiety, 4 = Interest in Foreign Languages, 5 = L2 Self Confidence, 6 = Instrumentality: Prevention, 7 = Instrumentality: Promotion, 8 = Attitude towards Learning, 9 = Intended Learning Effort, M = Mean Motivation

In summary, the quantitative analysis of the language level, LA and motivation of students enrolled in different CLIL classes revealed that science students in both 10th and 11th grade had a higher language level than economics students and outperformed them in the overall LAT and in four out of the five different prompts. However, significantly, no difference was found between the groups in the prompt *Economy & Money*, indicating that studying the content-related vocabulary in their economics class has allowed the economics students to bridge the gap between them and the science students in this particular lexical domain. Regarding motivation, no differences were found between the groups in 10th grade, and differences were found only in the categories *The “Ought to” Self* and *L2 Self Confidence* in 11th grade. Similarly, when focusing on the participants in the longitudinal analysis, no differences were found between the groups in either grade. This suggests that the differences between the groups in terms of language level and LA may not be attributed to language learning motivation.

8.4.2. Qualitative Differences in CLIL: Lexical Availability

Research question 4.2 asked whether there were qualitative differences in the words produced in English by students taking different CLIL classes at each testing period. In order to address this, the LA of science and economics groups in 10th and 11th grade, as outlined in the previous section, was analysed in terms of the frequency of first word responses for each content-related prompt, the most and least productive content-related prompts, and the lexical sophistication of each language based on the non-shared words of participants as well as the number of infrequent words in the production of each content-related prompt.

Firstly, analysis of the most frequent first words for content-relevant prompts by science and economics students (Table 8.78) showed that students in each group produced similar responses in both 10th and 11th grade. The only prompt which differed was *Economy & Money* in 10th grade, where science students’ most frequent first word was “economy” whereas economics students’ most frequent first word was “money”. These, notably, are two words which feature in the title of the prompt and so it is unsurprising that students would think of these first. Another observation was that in the economics group, there was more variety in the first word produced for this prompt. Other than the three students who produced the most common first word “money”, common first words included “risk” and “income”, which may show greater knowledge of the content-related vocabulary that economics students may be exposed to in their classes.

Table 8.78

Most Frequent First Words for Content-relevant Prompts by Science and Economics Students

10 th grade				
	Science (<i>n</i> = 26)		Economics (<i>n</i> = 15)	
Prompt	Lexical Unit	N of informants	Lexical Unit	N of informants
3	<i>Football</i>	8 (30.7%)	<i>Football</i>	6 (40%)
4	<i>Sun</i>	5 (29.2%)	<i>Soleil</i>	5 (33.3%)
5	<i>Economy</i>	8 (30.7%)	<i>Money</i>	3 (20%)
11 th grade				
	Science (<i>n</i> = 21)		Economics (<i>n</i> = 19)	
Prompt	Lexical Unit	N of informants	Lexical Unit	N of informants
3	<i>Football</i>	8 (38.1%)	<i>Football</i>	8 (42.1%)
4	<i>Sun</i>	5 (23.8%)	<i>Sun</i>	4 (21%)
5	<i>Money</i>	6 (28.5%)	<i>Money</i>	5 (26.3%)

Note. Prompt 3 = *Sport and Physical Activities*, Prompt 4 = *Environment & Climate*, Prompt 5 = *Economy & Money*.

Secondly, analysis the most and least productive content-related prompts (Table 8.79) again showed few differences between the groups. In both 10th and 11th grade, *Environment and Climate* was the most productive content-relevant prompt for both science and economics students. This was generally followed by *Sport & Physical Activities* and then *Economy & Money*, with just one exception: economics students in 10th grade produced more words for the prompt *Economy & Money* than for *Sport & Physical Activities*. This may indicate that, having studied content-related vocabulary in their economics class, these students were better able to produce vocabulary in this lexical domain. The same was not found for 11th grade economics students, perhaps due to the fact that a large number of students no longer studied economics through the medium of English.

Table 8.79

Ranking of Most and Least Productive Content-Relevant Prompts by Science and Economics Students

10 th grade				
Rank	Science	Tokens	Economics	Tokens
1	Environment & Climate	531	Environment & Climate	232
2	Sport & Physical Activities	478	Economy & Money	212
3	Economy & Money	360	Sport & Physical Activities	203
11 th grade				
Rank	Science	Tokens	Economics	Tokens
1	Environment & Climate	446	Environment & Climate	290
2	Sport & Physical Activities	405	Sport & Physical Activities	289
3	Economy & Money	296	Economy & Money	257

The analysis of students' lexical sophistication included an assessment of their non-shared words and the number of infrequent words in the production of each content-related prompt.

Regarding the non-shared words (Table 8.80), results again showed a great deal of similarity. The prompt *Economy & Money* contained the highest percentage of non-shared words in all groups. However, in the case of 11th grade students, the percentage for economics students was slightly higher than that of science students, indicating a slightly higher lexical sophistication in this measure. In 10th grade, the prompt *Environment & Climate* contained slightly more non-shared words for economics students (22%) than for science students (18%), whereas for 11th grade there was a similar number of non-shared words in the prompts *Environment & Climate* and *Sport & Physical Activities*.

Table 8.80*Non-shared Words in English by Science and Economics Students*

10 th grade				
Rank	Science	Non-shared words	Economics	Non-shared words
1	Economy & Money	109/360 30%	Economy & Money	64/212 30%
2	Sport & Physical Activities	96/478 20%	Environment & Climate	51/232 22%
3	Environment & Climate	96/531 18%	Sport & Physical Activities	43/203 21%
11 th grade				
Rank	Science	Non-shared words	Economics	Non-shared words
1	Economy & Money	95/296 32%	Economy & Money	92/257 36%
2	Environment & Climate	89/446 20%	Environment & Climate	65/290 22%
3	Sport & Physical Activities	70/405 17%	Sport & Physical Activities	52/289 18%

Finally, regarding the number of infrequent words in the production of each content-related prompt (Table 8.81), science and economics students exhibited similar profiles in 10th grade. One exception was in the prompt *Environment & Climate*, where science students produced relatively fewer words at the K1-K5 level and more off-list words, while the opposite was found for economics students. Greater variety was found in 11th grade, where differences were observed at the K1-K5 level and off-list level in all prompts. For the prompts *Sport & Physical Activities* and *Environment & Climate*, science students again produced fewer words at the K1-K5 level and more off-list words. However, interestingly, the reverse was true for the prompt *Economy & Money*, where economics students produced fewer words at the K1-K5 level and more in the off-list. This suggests that while 11th grade science students show higher lexical sophistication in terms of the number of infrequent words in the prompts *Sport & Physical Activities* and *Environment & Climate*, the economics students did so for the prompt *Economy & Money*. This finding is consistent with results of the quantitative analysis, which indicated that despite having a lower level of English and producing statistically fewer words than the science students in all other categories, the economics students did not produce statistically fewer words in the prompt *Economy & Money*. Thus, it appears that

economics students are not only improving their content-related vocabulary quantitatively, but also qualitatively. Given that the same results were not found qualitatively in 10th grade, it could be the case that more time is needed for students to improve their lexical sophistication than to increase the number of words produced.

Table 8.81

Frequency Distributions English by Science and Economics Students

10 th grade							
Prompt		K1-K5	K6-K10	K11-K15	K16-K20	K21-K25	Off-list
3	Science	67.8	10.7	1.4	0.4	0.2	19.46
	Economics	67.5	10.9	1.5	0	0.5	19.7
4	Science	84.8	4.9	0.2	0	0.4	9.79
	Economics	94.4	1.7	0.8	0	0	3.02
5	Science	82.2	6.7	0.6	0.3	0	10.28
	Economics	85.8	2.4	0	0	0	11.79
11 th grade							
Prompt		K1-K5	K6-K10	K11-K15	K16-K20	K21-K25	Off-list
3	Science	65.2	10.7	1.6	0.2	0	22.22
	Economics	70.2	9	1.6	0.7	0	18.34
4	Science	84.3	3.1	0.4	0	0.4	11.66
	Economics	92.8	3.1	0	0	0	4.14
5	Science	85.5	7.4	0.7	0.3	0	6.42
	Economics	75.1	9.3	1.9	0	0	13.62

In summary, the qualitative analysis has by and large shown rather similar trends in science and economics students in 10th and 11th grade, however, some interesting observations deserve attention. In 10th grade, differences were found in the most frequent first word for the prompt *Economy & Money* (“economy” for science students and “money” for economics students). In addition, more variety was observed by the economics students in the first word produced for this prompt. These same students also produced more words for the prompt *Economy & Money* than for *Sport & Physical Activities*, while the opposite was found for science students. Minor differences were also revealed in terms of non-shared words in 10th grade, where economics students produced

slightly more non-shared words for the prompt *Environment & Climate* than science students. Lastly, while no clear differences were found in the rest of the qualitative analysis, 11th grade students showed more differences in the analysis of the number of infrequent words produced, where economics students produced fewer words at the K1-K5 level and more off-list words in the prompt *Economy & Money*, while science students were found to do so in the other two prompts. While these results show only small differences between the groups at hand, they offer interesting observations which may indicate changes in progress in the qualitative nature of the students' lexical production.

8.4.3. Quantitative Differences in CLIL: Content-Relevant Prompts

Research question 4.3 aimed to determine if there was a difference in the number of words produced by students in English and French when the prompt was related to a CLIL class taken in either English or French. As outlined for RQ1.1⁴, paired samples t-tests were carried out on the three out of five prompts which were selected in order to tap into the content related vocabulary of the students CLIL classes: *Sports & Physical Activities* (Physical Education in English), *Environment & Climate* (Geography and History in French) and *Economy & Money* (Economics in English). As previously discussed, CLIL classes varied across the three grades under analysis: in 9th grade students studied physical education in English and geography and history in French but did not study economics; in 10th grade students studied Physical Education in English and geography and history in French and students studying economics did so in English; in 11th grade students no longer took CLIL classes with the exception of male students studying economics did, who did so through English. The results are shown in Table 8.82.

It was hypothesised that the prompts dealing with classes taken through English would yield a higher number of words in the English LAT, while the prompt dealing with the class taken through French would yield a higher number of words in the French LAT. However, as highlighted in Section 8.1, it was found that participants produced a higher number of words in the English LAT, both overall and in all five prompts. Thus, having studied geography through French has not led to students producing more words in French than in English in the prompt *Environment & Climate*. It is perhaps the case here that there is too large a difference in the language level of the students in English and French to see any noticeable difference in the prompts at hand.

⁴ See [Section 8.1.1](#) for a detailed overview of the descriptive statistics and normality tests.

Table 8.82*Differences in Lexical Availability in CLIL related Prompts*

9 th grade							
Prompt	M		SD		<i>t</i>	<i>df</i>	<i>p</i>
	English	French	English	French			
Sport & Physical Activities (EN)	14.68	9.82	5.96	4.02	5.52	37	<.001
Environment & Climate (FR)	17.31	10.21	7.63	5.28	6.72	38	<.001
Economy & Money	10.44	8.05	6.01	4.84	3.44	38	.001
10 th grade							
Prompt	M		SD		<i>t</i>	<i>df</i>	<i>p</i>
	English	French	English	French			
Sport & Physical Activities (EN)	16.61	11.22	4.56	2.56	7.74	40	<.001
Environment & Climate (FR)	18.61	10.41	5.86	4.75	10.67	40	<.001
Economy & Money (EN)	13.95	10.41	4.79	3.74	4.51	40	<.001
11 th grade							
Prompt	M		SD		<i>t</i>	<i>df</i>	<i>p</i>
	English	French	English	French			
Sport & Physical Activities	17.42	10.17	4.77	3.12	9.46	35	<.001
Environment & Climate	18.69	12.08	7.47	5.99	7.60	35	<.001
Economy & Money (EN-male)	13.94	9.47	5.11	4.52	5.56	35	<.001

Note. EN indicates that this prompt is related to a subject studied through English while FR indicates that this prompt is related to a subject studied through French.

Longitudinal analyses were also carried out to determine if there was a difference between the improvement in the content-relevant prompts which was dependent on the language of instruction, that is, did *Sport & Physical Activities* and *Economy & Money* improve to a greater extent in English than in French and did *Environment & Climate* improve to a greater extent in French. In order to analyse this, paired-samples t-tests were carried out, comparing the three content-relevant prompts at each data collection and in each language (Table 8.83).

Table 8.83*Longitudinal Differences in Content-relevant Prompts in English and French*

10 th → 11 th grade							
English (<i>n</i> = 32)							
Prompt	<i>M</i>		<i>SD</i>		<i>t</i>	<i>df</i>	<i>p</i>
	10 th	11 th	10 th	11 th			
Sport & Physical Activities	17.19	17.09	4.74	4.48	.152	31	.880
Environment & Climate	18.59	19.31	6.34	7.69	-.833	31	.411
Economy & Money	13.91	13.84	5.07	5.08	.098	31	.923
10 th → 11 th grade							
French (<i>n</i> = 30)							
Prompt	<i>M</i>		<i>SD</i>		<i>t</i>	<i>df</i>	<i>p</i>
	10 th	11 th	10 th	11 th			
Sport & Physical Activities	11.50	10.53	2.17	3.08	1.99	29	.056
Environment & Climate	10.87	13.13	4.47	5.77	-3.36	29	.002
Economy & Money	10.93	9.73	3.72	4.77	1.47	29	.152

Results indicated that there were no statistically significant differences between any of content-relevant prompts in the English LAT. However, with regards to French, statistically significant differences were found in the French content-relevant prompt *Environment & Climate*. Thus, it is possible that the exposure to French that 10th grade participants received in their geography and history CLIL class from the time of the first data collection until the end of the academic year was sufficient to see noticeable difference in this prompt the following year. The same was not found for the English content-relevant prompts. This is perhaps due to the high level of productivity in English, that is, given that participants already produced a particularly high number of tokens, there was less room for improvement, despite receiving exposure in their CLIL classes.

In summary, participants produced a higher number of words in the English LAT, both overall and in all five prompts, regardless of whether the prompt at hand was related to an English or French CLIL class. This can likely be attributed to the students' language level, which was much higher in English than in French. However, longitudinal analyses found that while there were no statistically significant differences in any of the content-relevant prompts in the English LAT from 10th to 11th grade, statistically significant

differences were found in the one French content-relevant prompt *Environment & Climate*. This finding is extremely important as it reveals that although studying geography through French does not result in the production of more tokens in the prompt *Environment & Climate* in French as compared to English, there is a clear advantage of being exposed to the content-related vocabulary, which can be seen in the longitudinal analysis.

8.4.4. Qualitative Differences in CLIL: Content-Relevant Prompts

Research question 4.4 aimed to determine whether there were qualitative differences in the words produced by students in English and French when the prompt was related to a CLIL class taken in either English or French at each testing period. In order to address this, a qualitative analysis was carried out to identify the frequency of first word responses for each content-relevant prompt, the most and least productive content-relevant prompts, and the lexical sophistication of each language based on the non-shared words of participants as well as the number of infrequent words in the production of each content-relevant prompt. This section provides a summary of the data provided in detail in Section 8.1.2, discussing the results with a focus on the content-relevant prompts so as to address research question 4.4.

Firstly, with regard to the most frequent first word for each prompt, results indicated that in each of the content-relevant prompts, the first word was remarkably similar in both languages across all three grades. In the prompt *Sport & Physical Activities*, “football” was the most common word across both languages at all levels, with the exception of 10th grade students in French, where “basketball” was the most common. In the prompt *Environment & Climate*, “sun” was the most common word across both languages at all levels. Greater differences were seen in the prompt *Economy & Money*, where for 9th grade students, it was “dollar” in English and “euro” in French, for 10th grade students it was “economy” in both languages, and for 11th grade students it was “money” in both languages. Thus, with the exception of “football” / “basketball” in *Sport & Physical Activities* in 10th grade and “dollar” / “euro” in *Economy & Money* in 9th grade, no differences were found across language in the content-relevant prompts in each language.

Secondly, prompts were analysed in order to determine the ranking of the most and least productive content-relevant prompts for both English and French in each grade. Notably, when focusing solely on the content-relevant prompts, results revealed that the

ranking was exactly the same in both languages for 9th and 11th grade students and in English for 10th grade students: the highest number of words were produced in the prompt *Environment & Climate*, followed by *Sport & Physical Activities* and then *Economy & Money*. In French in 10th grade, *Economy & Money* was again the prompt that produced the fewest words, but *Sport & Physical Activities* was found to be the most productive, followed by *Environment & Climate*. This, notably, is despite the fact that 10th grade studied geography through French and physical education through English. However, as previously discussed with regard to the level of cognateness, the prompt *Sport & Physical Activities* in French was found to contain the highest number of English cognates of all prompts (76% of the words produced were English cognates), which may account for the higher number of words produced in this prompt in French. Thus, again, there does not seem to be an impact of the language of instruction on the most and least productive content-relevant prompts in English and French.

Finally, lexical sophistication was investigated in terms of the non-shared words of participants as well as the number of infrequent words in the production of each prompt. Regarding non-shared words, each prompt was analysed in turn to determine the number of words which were unique to one participant, and the percentage of these non-shared words. Results again revealed remarkable similarities across both languages, with slight differences found again in 10th grade. In 9th grade, the percentage of non-shared words was exactly the same in each language, with *Economy & Money* producing the highest, followed by *Sport & Physical Activities* and then *Environment & Climate*. In 10th grade, *Economy & Money* had the highest number of non-shared words in both languages, though in English this was followed by *Sport & Physical Activities* and then *Environment & Climate* and in French it was followed by *Environment & Climate* and then *Sport & Physical Activities*. Thus, at this level, students produced a higher number of non-shared words in *Sport & Physical Activities* in English, which was related to physical education studied through English, and a higher number of non-shared words in *Environment & Climate* in French, which related to geography studied through French. In 11th grade, as in 9th grade, the percentage of non-shared words was exactly the same in each language. However, while *Economy & Money* again had the highest number of non-shared words, this was followed by *Environment & Climate* and then *Sport & Physical Activities*. Thus, with the minor exception of 10th grade students, who produced a higher number of non-shared words in English in the English content-relevant prompt *Sport & Physical Activities* and in French in the French content-relevant prompt *Environment & Climate*,

results generally indicated no difference in non-shared words dependant on the language of instruction.

The final part of the qualitative analysis consisted in an analysis of lexical sophistication in terms of the number of infrequent words produced in each language. As previously noted, given that the corpora used for each language are evidently different, it was not possible to directly compare the results between English and French. However, despite the differences in the corpora, clear similarities were found in both languages across all grades. In English, *Environment & Climate* contained the highest percentage of words in the K-1 to K-5 lists, followed by *Economy & Money* and then *Sport & Physical Activities*. The prompt *Sport & Physical Activities* also contained the highest number of off-list words, which as previously discussed can be attributed to the high number of multiword units within this category. Remarkably similar results were found in the French data: *Environment & Climate* contained the highest percentage of words in the K-1 to K-5 lists, followed by *Economy & Money* and then *Sport & Physical Activities*, at all levels. The percentage of words in the K-1 to K-5 lists was notably much lower than in English (around 35-39% in French as compared with 67-69% in English), though this could of course be attributable to the difference in the corpora. Again, the prompt *Sport & Physical Activities* also contained the highest number of off-list words, again attributable to the high number of multiword units within this category.

In conclusion, given the similarities in the English and French LA data, and in particular the similarities across the different grades at hand, it appears that the language of instruction made no major qualitative impact on the students' content-relevant prompts in English and French on the basis of the LATs.

Chapter 9: Discussion

This section provides a discussion of the results outlined in the previous chapter, addressing each area and research question in turn: Target Language Influence on LA, Target Language Influence on Language Learning Motivation, and the effect of Gender and CLIL on LA and motivation.

9.1. Target Language Influence on Lexical Availability

The first research question asked whether there were quantitative (RQ1.1) and qualitative (RQ1.2) differences in the LA of secondary school CLIL students in English as compared to French.

9.1.1. *Quantitative Differences in English and French LA*

Regarding the quantitative differences in the English and French LATs, it was hypothesised that, given the emphasis placed on learning EFL in Spain and the consequent higher number of hours of exposure that students generally receive, the participants would retrieve a higher number of words in English as compared with French. This was found to be the case: participants in 9th, 10th and 11th grade produced a statistically significant higher number of tokens in English than in French in all five prompts and overall. These results were also consistent with the participants' self-reported language levels, as students in all grades reported statistically significant higher language levels in English than French. While few previous studies have addressed the differences in LA in various FLs, these results are consistent with what has been found to date, which indicates that students generally produce a higher number of tokens in English as compared with other FLs. For example, Santos Díaz (2017c) found that participants produced the highest number of words in their native language, Spanish (397, 67 words), while those in the EFL group produced a mean of 261.67 in English and those in the FFL group produced a mean of 221.33 in French. Šifrar Kalan (2014) also found that for Slovenian university students, although clear similarities were found in the productivity of different prompts, the means for learners of English were slightly higher than for learners of Spanish. While these two studies compare different groups of students, the current study notably found similar results in participants studying both EFL and FFL simultaneously. In addition, the results of this study highlight the clear relationship between language level and LA. This is again consistent with suggestions in previous research that a higher number of words is produced by more advanced learners (van

Ginkel & van der Linden, 1996, as cited in Schmitt, 2000). Given that participants in this study reported higher language proficiency in English than in French, it is thus likely that the differences observed may not simply be ascribed solely to the participants' LA in each language, but rather to the language level that participants have in each one. It is thus suggested that, if we are to compare LA in English and other FLs, it is necessary to attempt to do so in learners whose language proficiency in the two TLs is more comparable. This would allow us to better understand whether differences in LA are solely attributable to language level or are a result of the specific language at hand.

Regarding the cross-sectional and longitudinal analyses, it was hypothesised that older, more advanced students would produce a higher number of tokens than younger students in each language, in keeping with findings from previous research (Agustín Llach & Fernández Fontecha, 2014; Jiménez Catalán & Fitzpatrick, 2014). However, the cross-sectional analysis revealed no statistically significant differences between 9th and 10th grade students in the overall LAT or the language level C-tests in either language. There was, however, a statistically significant difference in the prompt *Economy and Money* in English and in the prompts *Sport & Physical Activities* and *Economy & Money* in French. In all cases, 10th grade participants produced a higher number of tokens. For English, this finding is very likely related to that fact that 10th grade students studied economics through English, whereas the 9th grade students did not. For French, given that the prompts *Sport & Physical Activities* and *Economy & Money* were related to the CLIL classes Physical Education and economics, which were studied through English rather than French, it is possible that the 10th grade participants drew on their English lexical knowledge in order to perform better on these prompts. This is consistent with previous research which suggests that higher exposure to particular lexical domains results in higher productivity in the prompt at hand (Jiménez Catalán & Ojeda Alba, 2009a; Fernández Orío & Jiménez Catalán, 2015; Canga Alonso, 2017). Regarding longitudinal differences, results were highly dependent on the language at hand. In English, no statistically significant differences were found in the five prompts or the overall English LAT, although there was a statistically significant difference in the participants' performance on the English C-test. In French, however, statistically significant differences were found in two of the five prompts, *Animals* and *Environment & Climate*, as well as in the overall French LAT and in the French C-test. It should firstly be noted that, given that statistically significant differences were found in both the English and French C-tests, the participants have clearly improved their proficiency in both languages

over the course of the year. The differences in English and French in the LA from 10th to 11th grade can thus not solely be attributed to language level. However, it is suggested that, given the large gap between the number of tokens produced in English as compared with French, there was more room for improvement in French. In other words, there may be something of a ceiling effect at play, given the high number of tokens produced in English in 10th grade. Such results were found in previous productive vocabulary research by Alejo González and Piquer Píriz (2016a), who found statistically significant growth in the productive vocabulary of lower-level secondary school students, but not for higher-level students. Concerning previous LA research which has investigated age-related differences, upon comparing young learners with teenagers, Agustín Llach and Fernández Fontecha (2014) and Jiménez Catalán and Fitzpatrick (2014) found that older learners produced significantly more tokens than younger learners. While the former compared 6th grade and 9th grade students, the latter compared 6th grade and 8th grade students; the age gap is thus quite a bit larger than in the study at hand, where there is an age gap of just one year. It is interesting to note, however, that differences in LA are observed in the longitudinal analysis even after only one year, particularly in the case of French as opposed to English. These findings highlight the importance of the language at hand when carrying out longitudinal analyses, and in particular the potential differences in language level at the first data collection.

9.1.2. *Qualitative Differences in English and French LA*

Regarding the qualitative differences in the English and French LATs, this study aimed to determine whether there were differences in the words retrieved by participants in LATs in English as compared to French at each testing period with regards to a) the frequency of first word responses for each prompt, (b) the most and least productive prompts, and c) the lexical sophistication of each language based on the non-shared words of participants as well d) as the number of infrequent words in the production of each prompt. It was hypothesized that similarities would be observed in the frequency of first word responses for each prompt and the most and least productive prompts for general prompts (Santos Díaz, 2017c), but that when a prompt was related to a subject which was studied through the medium of one of the TLs, this prompt would be more productive in that language (Dalton-Puffer, 2008; Heras & Lasagabaster, 2015). In other words, the general prompts *Animals* and *Food & Drink* were expected to show similarities in the two languages, whereas the content-related prompts *Sport & Physical Activities*, *Environment*

& *Climate* and *Economy & Money* would show the positive effect of CLIL education on students' learning of content-related vocabulary, resulting in greater productivity when the prompt was related to a subject studied in the TL. Regarding lexical sophistication, it was expected that more non-shared words would be found in the prompts which were the least productive (Šifrar Kalan, 2014), that fewer non-shared words would be produced in English than in French (Santos Díaz, 2017c), and that participants would produce a higher percentage of infrequent words in the most productive prompt in their L2 than in their L3 (Jiménez Catalán & Fernández Fontecha, 2019). Finally, regarding qualitative differences at each testing period, it was hypothesised that the similarity or difference in the frequency of first word responses for each prompt and the most and least productive prompts at each grade would be prompt dependent. Specifically, the most and least productive prompts were expected to vary in the case of content-relevant prompts, when participants studied different subjects in each grade. It was also expected that older participants would have incorporated more words at the 1K and 2K levels, as opposed to off-list words, in each language, as in Jiménez Catalán and Fitzpatrick (2014).

Results firstly indicated that there was a large degree of similarity in the most frequent first words for each prompt in English and French. However, some differences were found in the prompt *Sport & Physical Activities* in 10th grade, where the most common first word was “football” in English but “basketball” in French, and in *Economy & Money* in 9th grade, where the most common first words was “dollar” in English but “euro” in French. These slight differences do not appear to be attributable to the participants' CLIL classes and the language they were taught in. Thus, the findings are consistent with research by Santos Díaz (2017c), who found similarities across languages in general prompts; however, there is no indication that the participants' CLIL classes have had a positive effect on students' learning of content-related vocabulary (Dalton-Puffer, 2008; Heras & Lasagabaster, 2015) in terms of the most frequent first word.

Secondly, much greater variety was observed in the ranking of the most and least productive prompt in English as compared to French. Particularly noteworthy was the fact that the English content-relevant prompts, *Sport & Physical Activities* and *Economy & Money*, were the least productive at all three grades in the English LAT, while the French content-relevant prompt, *Environment & Climate*, was found to be the most productive for 9th grade and 11th grade students in the French LAT. In other words, studying Physical Education and economics through English did not result in the production of more words for content-related prompts than other more productive ones

such as *Food & Drink* or *Animals* in the English LAT, whereas studying geography and history through French possibly provided 9th and 11th grade participants with enough lexical items to surpass the number of words produced for more productive prompts. This finding offers very interesting insights into suggestions in previous research that CLIL provides a positive learning environment for learning content-related vocabulary (Dalton-Puffer, 2008; Heras & Lasagabaster, 2015), as it appears that in terms of the most and least frequent words, this advantage is very much dependent on the language at hand. It remains to be seen whether this is a product of the specific languages, or whether it is again attributable to the ceiling effect, discussed above. In other words, it is possible that given their higher language proficiency in English, participants were better able to produce a higher number of tokens in the general prompts as opposed to the content-related ones, whereas given their lower proficiency in French, the additional exposure to content-related vocabulary made a much greater difference in the prompt *Environment & Climate*.

Thirdly, the analysis of non-shared words revealed evident similarities in the rankings in English and French: content-relevant prompts generally contained a higher number of non-shared words in both English and French, which is consistent with findings by Šifrar Kalan (2014), who found more non-shared words in the least productive prompts. However, participants generally produced a higher percentage of non-shared words in English than in French, with some minor exceptions: *Animals* and *Food & Drink* in 9th and 11th grade and *Environment & Climate* in 11th grade. This finding was unexpected as, curiously, the reverse was found in previous research by Santos Díaz (2017c), whose participants produced fewer non-shared words in English than in French. However, it is important to note that while the participants in Santos Díaz's were two different groups, one studying English and one studying French, the sample in this study consists of learners studying both languages. It is thus important to take into consideration the hierarchy in their language proficiency, which appears to play an important role. For example, more recent research by Agustín-Llach (2022) has revealed that older, more proficient learners, produce more idiosyncratic words, perhaps due to their wider vocabulary. It is thus likely here that more non-shared words were produced in English than in French as a result of the participants' higher language proficiency.

Finally, concerning the number of infrequent words produced in English and French, it should again be stressed that different corpora were necessarily used in the analysis of each language, making it impossible to directly compare the two languages.

However, similar findings were found in the case of each language. Notably, the less common semantic categories, *Environment & Climate* and *Economy & Money*, contained a higher percentage of words in the K-1 to K-5 lists than the other categories across all grades in both languages. This finding is partially consistent with findings by Jiménez Catalán and Fernández Fontecha (2019), who found that participants produced a higher percentage of infrequent words in the most productive prompt. In the current study, *Economy & Money* was found to be the least productive prompt at all grades in both languages, yet it contained a higher percentage of words in the K-1 to K-5 lists than other prompts. However, the French content-relevant prompt, *Environment & Climate*, was found to be the most productive for 9th grade and 11th grade students in the French LAT, and also contained a higher percentage of words in the K-1 to K-5 lists than other prompts. Thus, the hypothesis that a higher percentage of infrequent words would be found in the most productive prompts holds true for *Environment & Climate* in French, but not for *Economy & Money* in English and French. This implies that a higher number of words are produced in French in the prompt *Environment & Climate* by relying on more basic vocabulary, whereas in English and French, despite relying on this more basic vocabulary, the prompt *Economy & Money* remains one of the least productive. A high percentage of off-list words was also found in the prompt *Sport & Physical Activities* in both English and French and in the prompt *Food & Drink* in French, likely due to the presence of multiword units.

Regarding the cross-sectional and longitudinal qualitative analysis, despite the hypothesis that differences would be observed depending on the different CLIL classes taken at each grade, and that older participants would exhibit higher lexical sophistication, very few differences were observed. The most frequent first word for each prompt remained constant across grades in most prompts. One slight difference was found in 10th grade students for the prompt *Sport & Physical Activities* in French, where the most frequent first word was “basketball”, as opposed to “football”, at all grades for English and 9th and 11th grade for French. In addition, the prompt *Economy & Money* varied in 9th, 10th and 11th grade in both languages: respectively “dollar”, “economy” and “money” in English and “euro”, “economy” and “money” in French. However, despite these differences across the grades, there remains a clear reliance on currencies (“dollar” / “euro”) and vocabulary in the title of the prompt itself (“economy” / “money”). Thus, while there is evident variation across grades in this content-relevant prompt, this does not appear to be related to the differences in CLIL subjects, as was hypothesized.

Regarding the ranking of the most and least productive prompts, there was practically no differences across the grades for English, whereas some notable differences were found in 10th grade in French. While the content-related prompt *Environment & Climate* was the most productive prompt in 9th and 11th grade in French, it ranked third for 10th grade students. Given that 10th grade students, like 9th grade students, studied geography and history through French, it was expected that this prompt would also be the most productive, however, it was surpassed by *Sport & Physical Activities* and *Food & Drink*. Analysis of cognates across grades revealed a very similar pattern: practically no differences were observed for English, while some clearer differences were found in 10th grade in French. In English, the percentage generally did not differ by more than 2%, with the exception of *Economy & Money*, which had a very moderate difference of 7% between 9th and 10th grade. On the other hand, somewhat larger differences of up to 14% were found in French in most prompts, with the exception of *Food & Drink*, which again differed by only 2%. Notably, the highest percentages of cognates were found in the 9th grade students, who potentially relied more on similar English vocabulary to support their completion of the French LAT. As noted above, the content-relevant prompt *Environment & Climate* was found to be the most productive prompt in the French LAT for 9th and 11th grade students, but not for 10th grade students. For these participants, the prompt with the highest percentage of cognates, *Sport & Physical Activities*, was in fact that which ranked as the most productive prompt for French. This suggests that productivity here has been influenced by the presence of cognates, leading to differences in 10th grade as compared to 9th and 11th grade. Thus, although it was hypothesized that the most and least productive prompts would vary in the case of content-relevant prompts, when subjects studied differ in each grade, no differences were found across the three grades which could be attributable to their CLIL classes. Regarding lexical sophistication in terms of the ranking of non-shared words and the number of infrequent words in the production of each prompt, no real difference was observed across the three grades in either language. In the former, *Economy & Money* contained the highest and *Animals* contained the fewest number of non-shared words at all three grades, and content-relevant prompts also generally contained a higher number of non-shared words across all three groups in both languages. In the latter, no observable difference was found in the number of infrequent words across grades in either language. This is inconsistent with previous research by Jiménez Catalán and Fitzpatrick (2014), who found that older students incorporated more words at the 1K and 2K levels, as opposed to off-list words. This may be attributable to the age of the

students at hand, as in the previously mentioned research, participants were in 6th and 8th grade, whereas here we are dealing with older students in 9th, 10th and 11th grade. It is thus suggested that fewer differences are observable in terms of lexical sophistication when comparing older adolescents as opposed to comparing older children with younger adolescents.

9.2. Target Language Influence on Language Learning Motivation

The second research question asked whether there were quantitative differences between the participants' language learning motivation towards English as compared to French at each testing period (RQ2.1) and whether there was a relationship between the participants' LA, language level and their language learning motivation (RQ2.2) in each language at each testing period.

9.2.1. Quantitative Differences in Motivation towards English and French

Regarding the quantitative differences in the participants' language learning motivation towards each language, it was hypothesized that students at all levels would report higher language learning motivation towards English than to French, given previous research which has indicated that students report more positive attitudes and higher motivation towards English than to the other TLs (e.g., De Smet et al., 2018, 2019; Geoghegan, 2018). Results indicated that this was indeed the case: there were statistically significant differences in the participants' overall language learning motivation for each language, and nearly all categories, with participants' indicating a higher level of motivation towards English than to French in all cases and at all grades. This is consistent with findings in primary and secondary school students from De Smet et al. (2018), who found lower anxiety and higher enjoyment in English CLIL compared to Dutch CLIL, and De Smet et al. (2019), who found more positive attitudes and higher motivation in English CLIL compared to Dutch CLIL in terms of the categories *attractiveness*, *easiness*, *expectancy for success*, *perceived task value* and *cost*. While the cohort of students was rather different to the present study, Geoghegan (2018) used a similar instrument to investigate the L2MSS and found differences only in the category *Ideal L2 Self*. The results showed that Spanish university students focusing on English expressed a greater ability to visualize themselves as the L2 user they wished to be than those focusing on French or German. In the present study, however, even greater differences were observed overall and across the individual categories. This could well be due to the age group of the students, or the fact that the same cohort of students were compared, directly

comparing their motivation towards English and French.

Individual motivation categories which revealed no statistically significant differences included *The “Ought to” Self* in 9th grade, *Language Anxiety* and *L2 Self Confidence* in 10th grade, and *L2 Self Confidence* in 11th grade. The younger participants in the study thus indicated no key differences in the way they are externally motivated to learn English as opposed to French, while the older participants’ self-confidence and anxiety appeared to be similar regardless of the language at hand. Regarding the former, previous research has indicated that the *The “Ought to” Self* may be age-relevant, potentially becoming less important as students get older, with it playing a greater role in secondary school students, who often have no choice in whether or not they study the language at hand, than university students, who generally opt to study the language (Busse & Williams, 2010; Oakes, 2013). In addition, recent research has also suggested that this category may be even less relevant for LOTE learners (Oakes & Howard, 2019). This may be a factor here, as although French was compulsory in 9th and 10th grade for all participants, a number of the 11th grade students, namely the male students, could choose whether or not to continue with French in upper-secondary school. The latter finding is also particularly interesting as it indicates that if 10th and 11th grade students exhibit anxiety or lack confidence in one language, the same may also be true for the other. This was not the case in research by De Smet et al. (2018) who, as noted above, found that participants enrolled in English CLIL reported less anxiety than those in Dutch CLIL. However, this could be attributable to the fact that different groups of students were compared. The results of this study suggest instead that when the same students are compared, they tend to report similar levels of anxiety towards both TLs in terms of these affective factors.

Regarding the cross-sectional and longitudinal analysis, results revealed few differences from one grade to the next. In the cross-sectional analysis, statistically significant differences between students in 9th and 10th grade were found only in *The “Ought to” Self* in English, where 10th grade participants reported higher motivation towards external sources of motivation towards English. This could well be due to the fact that these students were in their final year of compulsory secondary education, and so may well have felt more pressure to perform better in English than the 9th grade students. In the longitudinal analysis, statistically significant differences from 10th to 11th grade were found only in *Attitude towards Learning* in English and *Instrumentality: Prevention* and *Instrumentality: Promotion* in French. The 11th grade students reported a

less positive attitude towards learning English and lower instrumentality towards learning French than they had in 10th grade. This could likely be due to the fact that, as 11th grade students generally no longer took CLIL classes in English, they may have experienced less enjoyment towards learning English. Regarding French, given that this language ceased to be a compulsory school subject for male students in 11th grade, a larger number of students may well have perceived the language to be less instrumentally important than they had in 10th grade. As there is a lack of previous research investigating differences in motivation towards learning English and LOTEs at different points in time, these results offer an interesting insight into how motivation towards English and other TLs develops over time. It would be particularly interesting if future research were to home in on the compulsory nature of learning the TLs at hand, and specifically compare two TLs which are both compulsory or are both not compulsory so as to determine how this plays a role in other cohorts of students.

9.2.2. Lexical Availability, Language Level and Language Learning Motivation

Regarding the relationship between participants' LA, language level and language learning motivation, it was hypothesized that students with greater language proficiency would report higher motivation in the CLIL context at hand (Navarro Pablo, 2018). It was expected that this would be particularly relevant in the case of French, due to previous suggestions in research that LOTE learners usually attain a higher language proficiency for highly specific and personalised reasons (Dörnyei & Al-Hoorie, 2017). In addition, it was posited that participants with higher language learning motivation would also produce a higher number of words of the LAT, as was found in Fernández Fontecha (2010).

Results indicated clear differences depending on the language at hand. In English, significant moderate positive correlations were found between language level and motivation in 10th and 11th grade, but not in 9th grade, and between LA and motivation at all levels. In French, no statistically significant correlations were found between language level and motivation at any level, while a significant moderate positive correlation was found between LA and motivation in 10th grade, but not in 9th or 11th grade. Firstly, regarding language level and motivation, suggestions from Navarro Pablo (2018) that students with a higher language level report higher motivation in a CLIL context appear to hold true here for older students in English but not for any students in French. This is all the more curious when taking into account the fact that 11th grade students generally

no longer took CLIL classes. In other words, while no relationship was found for 9th grade students, who were enrolled in CLIL classes, there was a relationship for 11th grade students, who were not. In addition, although Dörnyei and Al-Hoorie (2017) have highlighted the relationship between language proficiency and motivation in LOTEs, this was not found to be the case here: in no case was there a relationship between language proficiency and motivation in French. Secondly, regarding LA and motivation, while Fernández Fontecha (2010) found that a higher number of words in the LAT correlated with higher language learning motivation, in the present study this appears to be the case for all levels in English, but only in 10th grade for French. In 9th and 11th grade, students who reported higher French language learning motivation did not necessarily produce a higher number of tokens in the French LAT. These findings suggests that there are clear differences in the relationship between participants' LA, language level and language learning motivation depending on the language at hand, and while results for English are consistent with previous research, those for French are not: in English, students with higher language proficiency generally reported higher motivation and produced a higher number of tokens in the LA, whereas the same was not seen for French.

Regarding the individual categories, a number of interesting observations were made: no relationship was found between the LA in English and the *“Ought to” Self, Instrumentality: Promotion or Attitude towards Learning* at any level, while in French, no relationship was found between the LA and *Language Anxiety, Interest in Foreign Languages, Instrumentality: Prevention, Attitude towards Learning, or Intended Learning Effort*. Thus, English LA does not appear to be related to external sources of motivation, how English is seen to promote future success or the attitude towards the language. French LA, on the other hand, does not appear to be related to how anxious the students are, how interested they are in French, how not having French is seen as preventing future success, the attitude towards the language, or the amount of effort that is made to learn French. In general, there appears to be a greater relationship between LA and motivation in English as compared to French, though in no case is attitude towards language an important factor. Regarding *The “Ought to” Self* in English, similar results were found in a recent study by Sandu and Oxbrow (2021), who observed that while a stronger *Ideal L2 Self* correlated with a wider LA, *The “Ought to” Self* showed marginal relevance. However, the same study also found a relationship between higher productivity and a more positive *L2 Learning Experience*, whereas in the current study no correlation was found between LA and *Attitude towards Learning* in either language. This could be

due to the fact that the participants in Sandu and Oxbrow (2021) were university students rather than secondary school students, which likely made a key difference in their attitudes toward learning.

9.3. Gender

The third research question asked whether there were quantitative differences in the words retrieved by male and female students in LATs in English and French at each testing period, both across languages (RQ3.1.1) and within languages (RQ3.1.2), whether there were qualitative differences in the words retrieved by male and female students in a LAT in English as compared to French at each testing period (RQ3.2), and whether there were quantitative differences between male and female students with regards to language learning motivation in English and French at each testing period, again both across languages (RQ3.3.1) and within languages (RQ3.3.2).

9.3.1. *Quantitative Differences in Gender and LA*

Firstly, regarding the quantitative differences in the words retrieved by male and female students in the LATs in English and French at each testing period, it was hypothesized that both male and female students would produce a higher number of words in English as compared to French, as in RQ1, and that female students would produce a higher number of words than male students, in keeping with previous research (Jiménez Catalán & Ojeda Alba, 2009a; Fernández Fontecha, 2010; Agustín Llach & Fernández Fontecha, 2014). However, these differences were expected to become less pronounced as students got older, given findings by Jiménez Catalán and Canga Alonso (2019) which found no gender difference in older 12th grade Spanish students.

9.3.1.1. Quantitative Differences in Gender and LA across Languages. With regard to the quantitative differences in gender and LA across languages (RQ3.1.1), results indicated that both male and female students demonstrated a higher productive vocabulary in English as compared with French, which is consistent with the findings for RQ1. This indicates that, regardless of gender, students have a higher productive vocabulary in English as compared to French.

However, when taking into consideration the different testing periods, some interesting gender-based differences were observed regarding time and the interaction between time and language. Regarding time, results of the cross-sectional analysis showed that for male participants there was significant main effect of time on the C-tests,

the prompt *Food & Drink* and the prompt *Economy & Money*, with higher means in 10th grade, whereas for female participants there was significant main effect of time only on the prompt *Animals*, with higher means in 9th grade. This indicates that there was a greater different difference between males in 9th and 10th grade than females in 9th and 10th grade. In the former, 10th grade students' language level was significantly better in both languages, and they produced a significantly higher number of tokens for *Food & Drink* and *Economy & Money*. Notably, *Economy & Money* was related to the CLIL subject economics which students began to study in 10th grade. It thus makes sense that these students would produce a higher number of words in this lexical domain. As for the latter, it is surprising that 9th grade female students in fact produced more words for *Animals* in the two languages, as it would be expected that the older, 10th grade students would produce a higher number of words. This could potentially be attributable to group differences. Results of the longitudinal analysis indicated that for male participants there was no significant main effect of time on the C-tests, the overall LAT or any of the five individual prompts, whereas for female participants there was a significant main effect of time on the overall LAT and the individual prompts *Animals*, *Food & Drink*, and *Environment & Climate*, with statistically significant higher means in 11th grade. It thus appears that female students improved to a much greater degree than male students both in their overall LA and specifically within these lexical domains across the two languages. This difference between the cross-sectional results, where male students showed greater improvement, and the longitudinal results, where female students showed greater improvement may evidently be due to the differences in the nature of the collection. It would be extremely beneficial for future research to carry out further longitudinal research across a longer period of time, for example, across three grades, in order to confirm whether the results found here are due to the grade the students are in or whether they can be attributed to differences in the cohorts of students.

Regarding the interaction between time and language, results of the cross-sectional analysis revealed that for male participants there were no statistically significant interactions between time and language in the C-tests, the overall LAT or any of the five individual prompts, whereas for female students, statistically significant interactions were found in the individual prompts *Sport & Physical Activity* and *Economy & Money*. From 9th to 10th grade, means for *Sport & Physical Activity* increased in English and decreased in French while for *Economy & Money* they increased in English and remained largely the same in French. This is likely attributable to the increased exposure to content-related

vocabulary in the class Physical Education, which was taught through English in 9th and 10th grade, and economics, which was taught through English in 10th grade. This is consistent with the suggestion that CLIL instruction improves content-related vocabulary (Dalton-Puffer, 2008; Heras & Lasagabaster, 2015); however, here it appears that this is the case only for female students. Results of the longitudinal analysis showed that there was a statistically significant interaction between time and language for male participants in the prompt *Economy & Money* and for female participants in the prompt *Food & Drink*. For male participants, the prompt *Economy & Money* increased in English from 10th to 11th grade but decreased in French. This again may be attributed to exposure to content-related vocabulary in the students' economics class. Here, it should be stressed that while male students continued to study this subject through the medium of English, female students did not. It is thus suggested that this extra exposure is what led to this significant interaction for male students. On the other hand, for female participants the prompt *Food & Drink* increased greatly in English and only slightly in French from 10th to 11th grade, indicating a much larger improvement in this general lexical domain in English as compared to French.

In sum, both male and female students were found to produce a higher number of words in English than in French, which is consistent with the results of RQ1 and indicates that these results are not influenced by gender. Analysis of the results at each testing period does, however, point to some interesting differences in how male and female students progress in the two languages. Of particular note are those which are related to CLIL classes, for example, higher means across the two languages for *Economy & Money* in males for 10th grade as compared to 9th grade males; an increase in English for *Sport & Physical Activity* and *Economy & Money* from 9th to 10th grade for female students; and an increase for *Economy & Money* in English from 10th to 11th grade for male students. As noted above, these changes are likely due to the differences in CLIL vocabulary exposure, which has been highlighted as beneficial (Dalton-Puffer, 2008; Heras & Lasagabaster, 2015), and so will be addressed in more detail in Section 9.4.

9.3.1.2. Quantitative Differences in Gender and LA within Languages. With regard to the quantitative differences in gender and LA within languages (RQ3.1.2), results showed that female students produced a higher number of words than male students in the overall LAT in both languages across all three grades with just one exception: the French LAT in 10th grade. An advantage was also seen in the individual prompts: females

produced a higher number of words in both languages and at all grades for the prompt *Environment & Climate* and across all prompts in French in 9th grade. This finding is consistent with previous research which has found female students to outperform male students in LAT tasks in English (Jiménez Catalán & Ojeda Alba, 2009a; Fernández Fontecha, 2010; Agustín Llach & Fernández Fontecha, 2014). Notably, these differences were found even in the older group, and so it appears that, at least until 11th grade, there is a clear female advantage. Thus, in this study, these differences did not become less pronounced as students got older, as was hypothesized based on findings from Jiménez Catalán and Canga Alonso (2019) in older 12th grade students. However, it should be noted that in the 11th grade group, female participants outperformed male students in all prompts of the English LAT with the exception of just one prompt: *Economy & Money*. This finding is particularly important, given the fact that while male students continued to study economics through English in 11th grade, female students did not. It is thus possible that having continued to receive exposure to content-related vocabulary in English allowed 11th grade male students to bridge the gap between them and their female peers (see also [Section 9.4](#)).

Regarding the performance at difference testing periods, the cross-sectional analysis revealed practically no similarity in the way male and female participants progressed, with no statistically significant difference from one grade to the next in both male and female students. In other words, any changes from one grade to the next were unique to one gender: 10th grade male students produced more words than 9th grade male students in the prompt *Food & Drink* in English and in the prompts *Sport & Physical Activities* and *Economy & Money* and the overall LAT in French, whereas 9th grade female students produced more words than 10th grade female students in the prompt *Animals* in French and 10th grade female students performed better than 9th grade female students in the prompt *Economy & Money* in English. These findings indicate that, although various increases are often observed from one grade to the next, male and female students' improvements appear to vary in different prompts and languages. In English, older male students performed better in *Food & Drink*, whereas female students performed better in *Economy & Money*. In French, older male students performed better in *Sport & Physical Activities* and *Economy & Money* and the overall LAT, whereas older females did not perform significantly better than the younger female group. On the contrary, 9th grade female students produced a statistically significant higher number of words for the prompt *Animals*. Another important observation was that older male

students performed significantly better on the language level C-tests, while the same was not found for female students. Thus, the larger differences seen in the case of the French LAT for male students are likely due to the improvement in language level. The longitudinal analysis also highlighted some interesting differences between male and female students. Results showed a significant main effect of gender on the C-tests, the overall LAT and all prompts except *Animals* in English, and a significant main effect of gender only on the overall LAT and the prompt *Environment & Climate* in French. This points to much larger gender differences in English than in French, with female participants outperforming males across the two data collections. These results are surprising, given that both genders received the same exposure to each language in 10th grade and reduced their exposure to both languages in 11th grade, although male economics students did continue to study this subject through English and so consequently received slightly more exposure than female students. However, it could be the case that, as students generally received more instruction in English, this led to a greater difference in this language between the two genders starting in secondary school that is, female students responded to the higher amount of exposure in a way that male students did not. Regarding time, statistically significant differences were found in the C-test and the prompt *Animals* in English, and in the C-test and the prompts *Animals*, *Environment & Climate* and *Economy & Money* in French. Thus, both males and females' LA improved to a greater extent in French than English, possibly due to the fact that their English language level was much higher than their French language level and so consequently there was less room for improvement. This was notably found to be the case in RQ1.1 and is consistent with previous research by Alejo González and Piquer Píriz (2016a) who found productive vocabulary growth only in lower-level secondary school students, but not higher-level students. Regarding the interaction between gender and time, there was a statistically significant interaction only in the prompt *Food & Drink* in English, whereas there was a statistically significant interaction between gender and time in the overall LAT and the prompt *Economy & Money* in French. In both English and French, male students' score decreased from 10th grade to 11th grade, whereas females' scores increased from 10th to 11th grade. There was consequently a much greater difference between male and female participants in 11th grade than in 10th grade. These results imply that not only did the female participants generally produce a higher number of words in each grade, they also improved to a greater extent from 10th to 11th grade. This finding is again inconsistent with results by Jiménez Catalán and Canga Alonso

(2019), who found no gender difference in the LA older 12th grade Spanish students, as in the current study female students appear to continue to improve to a greater extent than their male peers.

9.3.2. *Qualitative Differences in Gender and LA*

Secondly, regarding the qualitative differences in the words retrieved by male and female students in the English and French LATs at each testing period, previous research by Agustín Llach and Fernández Fontecha (2014) indicated similarities across genders in the most and least productive prompts, and so this was expected in the current study. However, it was hypothesised that when male and female students studied different content subjects in the TLs, those who received more prompt-related vocabulary exposure would produce a higher number of words in that category.

Regarding the frequency of first word responses, results revealed no gender differences in the prompts *Animals, Food & Drink, Environment & Climate* in either language, though some minor differences were observed for *Sport & Environment* and *Economy & Money*. In the former, male students' most frequent first word in both languages was "football", while females' most frequent first word varied to a greater extent in language and grade: in 9th, 10th and 11th grade, respectively, in English it was "tennis", "basketball" and "basketball" / "football" while in French it was "swimming", "basketball" and "basketball". It is suggested that these differences are related to student interest, given the fact that football is often more popular among male students and basketball is often more popular among female students (Tuero, González-Boto & Zapico, 2014). In the latter, although differences were observed, the most frequent words were generally found in the prompt name itself ("economy" / "money") or were a type of currency ("euro" / "dollar").

Regarding the most and least productive prompts, results found great similarity across almost all grades in English and in 9th grade in French, with *Food & Drink* being the most productive and *Economy & Money* being the least productive. This is consistent with previous research on Spanish EFL learners which found no difference in the most and least productive fields for male and female learners (Agustín Llach & Fernández Fontecha, 2014). In French, while the prompt *Sport & Physical Activities* ranked first for male students in both 10th and 11th grade, it ranked third in 10th grade and last in 11th grade for female students. The prompt *Environment & Climate* ranked second in 10th grade and first in 11th grade for female, whereas it ranked last in 10th grade and third in 11th grade

for male students. Thus, *Sport & Physical Activities* appears to be a more productive prompt for older male students and *Environment & Climate* appears to be a more productive prompt for older female students in French. These findings may possibly be related to student interest. Recent research by Lee and Pulido (2017), which investigated the effect of topic interest, language proficiency and gender on L2 vocabulary acquisition in reading, revealed a significant effect of topic interest and L2 proficiency, as well as a significant relationship between topic interest and gender for word-form recognition. The results highlighted how interest was more important for male students than for female students, as while females recognised significantly more word forms than males in lower interest texts, males recognised significantly more word forms than females in higher interest texts. Thus, it may be the case that, given greater interest in the subject, males may consequently have retained a higher number of words for the prompt *Sport & Physical Activities*, even in French.

Regarding lexical sophistication based on the non-shared words, male and female students' ranking was very similar at each grade in English, though in almost all cases male students had a higher percentage of non-shared words, suggesting a higher level of lexical sophistication. In French, 9th and 11th male students again had a higher percentage of non-shared words than female students in almost all cases, while 10th grade female students produced a higher number of non-shared words than male students across all prompts.

Regarding lexical sophistication based on the number of infrequent words in the production of each prompt, few differences were found between male and female participants in each grade. However, it was observed that female participants often produced a higher percentage of words at the K1-K5 band, while male students produced a higher percentage of off-list words. In other words, while the quantitative analysis revealed that female students produce a higher number of words than male students, this may be due to the fact they produce simpler, more common vocabulary. On the other hand, male students may produce fewer words, but the words they produce may be more lexically sophisticated. This is consistent with recent research by Li, Chen and Banerjee (2020) on writing task test scores, which indicated that, although the difference was not statistically significant, female test takers showed marginally lower lexical sophistication in their writing. Similar results were also found by Tankó (2021) who found evidence of greater lexical sophistication and varied lexis in the argumentative essays of male writers.

In sum, the above findings reveal few differences in the qualitative nature of male

and female students' LA in English and French. However, some small variations offer interesting insights. For example, the analysis of the most frequent first word may potentially be related to student interest, with male students most commonly producing "football" first as opposed to "basketball" by female students. The analysis also suggested that male students, although they produced fewer words than female students, may demonstrate greater lexical sophistication based on their higher percentage of non-shared words and higher percentage of off-list words, which is consistent with recent research on writing. This finding deserves further research in the field of LA, as it may partially explain some of the quantitative findings discussed above, that is, male students may produce fewer words in the LAT, but perhaps this is because they spend more time producing words which are more sophisticated. This finding is consistent with recent LA research by Agustín-Llach (2019) on bilingual learners, which compared students studying English as an L2 and those studying it as an L3. In this study, quantitative analysis showed that the monolingual L2 learners produced a higher number of words than the L3 learners, while qualitative analysis revealed that the bilingual L3 group tended to produce more words from lower frequency levels and off-list words. Thus, the monolingual group produced more words, but fewer lexical sophisticated words, that is, the quantitative advantage may come from producing simpler vocabulary. In the present study, male participants behave in a similar way to the L3 learners in Agustín-Llach (2019): they produce fewer words, but the words that they do produce appear to be more lexically sophisticated. This pattern offers interesting insight into LA analysis and deserves further attention in future research.

9.3.3. Quantitative Differences in Gender and Motivation

Finally, regarding language learning motivation, it was hypothesised that both male and female students would report higher motivation towards English as compared to French, as in RQ1, and that female students would report higher motivation towards each language, in keeping with previous findings in the field (Fernández Fontecha, 2010; Fernández Fontecha, 2014c; Lasagabaster & Sierra, 2009; Lasagabaster, 2016; Merisuo-Storm, 2006; Sylvén & Thompson, 2015; Calafato & Tang, 2019). However, it was also expected that CLIL instruction may lead to a blurring effect on motivation, as suggested by Gallardo-del-Puerto and Blanco Suárez (2021), given the theory that male students may be more motivated in a CLIL context due to greater interest in the subject (Heras & Lasagabaster, 2015). To this effect, it was expected that in cases where male students

expressed greater interest in their content classes, they would consequently have greater language learning motivation.

9.3.3.1. Quantitative Differences in Gender and Motivation across Languages. With regard to the quantitative differences in gender and motivation across languages (RQ3.3.1), results revealed that, in the case of both male students and female students, there were statistically significant differences between participants' motivation towards English and French in all grades for overall motivation, and in almost all individual categories. Some exceptions for male participants included *The "Ought to" Self* in 9th grade; *Language Anxiety*, *L2 Self Confidence*, and *Instrumentality: Prevention* in 10th grade; and *Language Anxiety*, *Interest in Foreign Languages* and *L2 Self Confidence* in 11th grade; while exceptions for female participants included *The "Ought to" Self* and *Language Anxiety* in 9th grade; *L2 Self Confidence* in 10th grade; and *Interest in Foreign Languages*, *L2 Self Confidence* and *Attitude Towards Learning* in 11th grade. Notably, no differences were observed for either male or female participants for the categories *The Ought to" Self* in 9th grade, *L2 Self Confidence* in 10th grade, and *Interest in Foreign Languages* and *L2 Self Confidence* in 11th grade. This indicates that, regardless of gender, there is little difference in how externally motivated participants are to each language in 9th grade, how confident they are in each language in 10th grade, and how confident and interested they are in each language in 11th grade. However, what is of particular interest here are the differences that arise, namely, in *Language Anxiety* in 9th grade; *Language Anxiety*, and *Instrumentality: Prevention* in 10th grade; and *Language Anxiety* and *Attitude Towards Learning* in 11th grade. Firstly, statistically significant differences were found for *Language Anxiety* for male learners in 9th grade but not for female learners. This suggests that while male students report lower anxiety for English than for French, female learners report similar levels of anxiety for both languages. This finding appears to be related to the students' self-perceived proficiency level: while there was a statistically significant difference between self-reported language level in English and French for male students, with students reporting a higher level in English, there was no such difference for female students. Thus, male students, who appear to view a greater difference between their proficiency in each language, report lower anxiety in English, in which they see themselves as more proficient. Female students, however, see less of a difference between their language proficiency in each language and consequently report similar anxiety toward each one. Secondly, statistically significant differences were found

in 10th grade for *Language Anxiety* and *Instrumentality: Prevention* for female learners, but not for male learners. This indicates that while female students report lower anxiety for English than for French, male learners report similar levels of anxiety for both languages. In addition, female students report that not having English will prevent their future success to a greater extent than not having French, whereas for male learners there is no difference between the two languages. As in the case of 9th grade, these results appear to be related to self-perceived language proficiency: no statistically significant difference was found between self-reported language level in English and French for male students, while female students reported a statistically significant higher level in English than in French. Finally, statistically significant differences were again found in 11th grade for *Language Anxiety* for female learners but not for male learners, and for *Attitude Towards Learning* for male learners but not for female learners. This suggests that while female students report lower anxiety for English than for French, male learners report similar levels of anxiety for both languages, and while male participants expressed a more positive attitude towards learning English than French, female participants' attitude towards the two languages showed no statistically significant difference. As in 10th grade, no statistically significant difference was found between self-reported language level in English and French for male students, while female students reported a statistically significant higher level in English than in French. Thus, it again appears that students' anxiety is related to their self-reported language proficiency in each language. This relationship between self-reported language proficiency and language anxiety has also been observed by Iqbal and Yongbing (2018) in Pakistani EFL students, who also found that female students reported higher level of L2 anxiety than male students. However, while in the present study *Language Anxiety* revealed differences at all grades, in 9th grade it was male students who reported lower anxiety for English than for French, whereas in 10th and 11th grade, it was female students report lower anxiety for English than for French. One reason for this could be the different cohorts of students. In other words, given that the group in 9th grade was different to the group in 10th and 11th grade, it could simply be the fact that this particular group of 9th grade male students were less anxious towards English than their other L2. However, it could also be the case that age is a factor; as is postulated by Zhang (2019), the role of anxiety may become more profound as age increases and findings of a metaregression indicate that there is a higher anxiety-performance correlation in older adult learners. Thus, it could be the case here that the older students became increasingly anxious towards learning English, leading to similar

levels of anxiety towards each language in 10th and 11th grade. This finding deserves future attention in research so as to determine the cause of the differences across grades in these adolescent learners. Another noteworthy finding is the differences observed for *Attitude Towards Learning* for male learners, who expressed a more positive attitude towards learning English than French in 11th grade. It is suggested that these differences are due to the fact that in 11th grade male students no longer studied geography in French, but students enrolled in economics continued to do so through English. Thus, if, as Heras and Lasagabaster (2015) suggest, male students are more motivated in a CLIL context due to greater interest in the subject, these students may have reported a more positive attitude towards English due to the continuation of CLIL education on the part of the economics students.

Concerning the cross-sectional analysis, the results indicated that there was a clear difference between 9th grade and 10th grade students depending on gender and language. While male participants' motivation towards English was relatively similar in 9th and 10th grade, older 10th grade students reported higher motivation towards French than the 9th grade students both overall and in a number of categories, such as *The "Ought to" Self*, *Interest in Foreign Languages*, *Instrumentality: Prevention* and *Instrumentality: Promotion*. The same, however, was not found for female participants, who reported relatively similar motivation towards each language in both grades. One exception to this was in the category *The "Ought to" Self* in English, where 10th grade female students reported higher external motivation. Curiously, while there are differences in *The "Ought to" Self* between 9th and 10th grade for both male and female students, for male students this difference arises only in French and for female students it arises only for English. This implies that while the older male students are more externally motivated towards French than the younger male students, the older female students are more externally motivated towards English than the younger female students. Of the most interest in these findings, however, is that fact that while older male students reported higher overall motivation towards French than their younger peers, the same was not found for female learners. While the students at hand all received the same amount of instruction in English and French, they did attend two different schools and so consequently had different teachers. It is thus possible that in the male school 10th grade students received greater pressure to perform well in French than they had in 9th grade, while in the case of the female school a similar situation arises in the students' English classes, resulting in the differences in *The "Ought to" Self*. In addition, given that these results come from a cross-

sectional analysis, it is of course possible that the differences are due to the fact that different cohorts of students are being compared. It would thus be beneficial to address these issues longitudinally from 9th to 10th grade. Concerning the longitudinal analysis, differences were again observed between language learning motivation towards English and French depending on gender. Male students' language learning motivation evolved in the same way regardless of language. In both English and French, 11th grade students reported lower motivation overall and lower motivation due to external sources than they had in 10th grade. Meanwhile, female students' motivation also evolved in a similar way in each language, though they reported a less positive attitude towards learning English in 11th grade than they had in 10th grade, while the same was not found in French. In sum, results of the cross-sectional analysis indicated a great deal of variation in motivation towards each language from 9th to 10th grade depending on gender, with 10th male students reporting much higher motivation towards French. However, given the lack of differences found in the longitudinal analysis, it is suggested that these differences may arise due to the differences in the groups in the cross-sectional analysis, rather than a difference that can be attributed to age. In other words, it appears that motivation towards English and French generally develops from one grade to the next in a similar way regardless of gender.

9.3.3.2. Quantitative Differences in Gender and Motivation within Languages. With regard to the quantitative differences in gender and motivation within languages (RQ3.3.2), results varied greatly depending on the language under analysis. In English, there were statistically significant differences only in the category *Instrumentality: Prevention* in 9th grade, and no statistically significant differences in 10th or 11th grade. In this case, 9th grade female students reported that they saw not having English as preventing their future success to a greater degree than the 9th grade male students. However, it is clear that, in general, both male and female students are equally motivated towards learning English. These findings are not at all consistent with previous research investigating motivation gender differences, which has by and large found female students to be more motivated than their male peers (Merisuo-Storm, 2006; Lasagabaster & Sierra, 2009; Fernández Fontecha, 2010; Fernández Fontecha, 2014c; Sylvén & Thompson, 2015; Lasagabaster, 2016; Calafato & Tang, 2019).

In French, much clearer differences were found in 9th and 10th grade: statistically significant differences were found in the categories *Instrumentality: Prevention* and

Instrumentality: Promotion in 9th grade, and in the categories *Language Anxiety*, *Instrumentality: Prevention* and *Attitude towards Learning* and the overall motivation in 10th grade. As in the case of English, 9th grade female students saw not having French as preventing their future success to a greater extent than the male students, whereas in 10th grade male students saw this to be more important than 10th grade female students. Female students in 9th grade also saw French as more important in promoting their future success than male students, whereas male students in 10th grade reported lower anxiety and a better attitude towards learning than female students. Notably, as in the analysis of motivation across languages, here there are clear differences between the students in 9th and 10th grade: while in the former female students are more concerned about the effect of not having French, in the latter this is more important to male students. Again, it is unclear here whether these differences may be attributed to the different cohort of students or another factor such as age or language exposure. Given that no differences were observed in 11th grade students, who were largely the same group of students as 10th grade, it is again suggested that differences in anxiety may be attributed to age.

Regarding *Attitude towards Learning* and the hypothesis that when male students expressed greater interest in their content classes, they would consequently have greater language learning motivation, analysis of the individual motivation questions offered some particularly interesting insights. In English, 9th grade male students reported higher interest in learning biology and geology in English, 10th male students reported higher interest in learning economics in English, and 11th grade students, reflecting on studying physical education in secondary school, reported higher interest in this subject and reported greater enjoyment as compared to female students. In French, both 9th and 10th grade male students reported higher enjoyment when studying geography and history than female students, and 11th grade students, reflecting on studying geography and history through French in secondary school, reported higher interest than female students. These results should clearly be interpreted with caution, as the study included only two questions tapping into each CLIL subject. However, the results do point to gender differences in how interested students are towards their CLIL subjects and how enjoyable they find them. This is consistent with the hypothesis that CLIL instruction may lead a blurring effect on motivation (Gallardo-del-Puerto & Blanco Suárez, 2021) due to greater interest in the subject (Heras & Lasagabaster, 2015). It would thus be extremely interesting for future research to incorporate further questions addressing each specific CLIL subject, in order to increase the internal reliability of the scales.

Concerning the cross-sectional analysis, there were evident differences between male and female students' language learning motivation depending on the language at hand. While there were few differences in English, clear differences arose in French between 9th and 10th grade male students, but not between 9th and 10th grade female students. Specifically, as compared to 9th grade male students, 10th grade male students reported higher motivation due to external sources, greater interest in learning French, saw French as more important in promoting their future success, and saw not having French as more preventative to their future success. However, female participants reported similar motivation towards French in both 9th and 10th grade. These findings are consistent with those discussed above. As noted, in the category *Instrumentality: Prevention* 9th grade female students reported higher motivation than the 9th grade male students, while 10th grade male students reported higher motivation than 10th grade female students. It is thus understandable that there was less of a difference between 9th and 10th female students, as the younger students already had high levels of motivation, and more of a difference between 9th and 10th male students, as the older levels had higher levels of motivation. Concerning the longitudinal analysis, results revealed greater differences from one grade to the next for male students than for female students. While this was observed in fewer categories, it was found for both English and French. Specifically, 11th grade male students reported lower overall motivation and lower motivation due to external sources than 10th grade male students in both languages, while female learners reported a less positive attitude towards learning English in 11th grade than in 10th grade. It is again suggested that this decrease in motivation may be attributed to a reduction in CLIL teaching in 11th grade. Notably, this appears to be particularly important for male students, perhaps due to suggestions that CLIL teaching plays a greater role in increasing male students' motivation towards the TL, as discussed above. The role of CLIL and the effect that it had will be discussed in greater detail the following section.

9.4. Content and Language Integrated Learning

The fourth research question addressed the effect of students taking different CLIL classes in English at each testing period (science as compared to economics), as well as the LA differences when a prompt was related to a CLIL class taken in either English or French at each testing period. It firstly asked whether there were quantitative differences in participants' language level, LA and motivation (RQ4.1) and whether there were qualitative differences in the words produced in English of students taking different

CLIL classes at each testing period (RQ4.2). It secondly asked whether there were quantitative (RQ4.3) and qualitative differences (RQ4.4) in the words produced by students in English and French when the prompt was related to a CLIL class taken in either English or French at each testing period. Regarding the former, this study aimed to answer calls by Canga Alonso (2017) for including LA prompts which are relevant to the CLIL subject alongside a measure of language proficiency, so as to determine whether there is a relationship between language level and LA in content-specific prompts. It was expected that when taking the content-related prompts into consideration, differences would arise in LA, both quantitatively and qualitatively, depending on the exposure students have received to this vocabulary in their CLIL classes, due to the purported positive effect that CLIL has on students' content-related vocabulary (Dalton-Puffer, 2008; Heras & Lasagabaster, 2015). It was also hypothesised that participants who expressed higher motivation towards a given CLIL subject would produce a higher number of words in the related prompt. Regarding the latter, based on research on secondary Spanish CLIL which found CLIL groups to outperform non-CLIL groups (Jiménez Catalán & Agustín Llach, 2017), it was hypothesised that prompts related to CLIL classes in English would result in a greater number of words in the English LAT, while prompts related to French would result in a greater number of words in the French LAT. In addition, qualitative differences were also expected, in particular regarding lexical sophistication in prompts which are related to their CLIL classes in each language.

9.4.1. Quantitative Differences in CLIL: Language Level, Lexical Availability and Motivation

With regard to the quantitative differences in participants' language level and LA, results indicated that in both 10th and 11th grade, science students had a higher language level than economics students and outperformed them in the overall LAT and in four out of the five different prompts. No statistically significant difference was found between the two groups in the prompt *Economy & Money*. This finding is incredibly important, as although the science students had a higher language level and produced a statistically significant higher number of tokens overall and in all other prompts, having studied the content-related vocabulary in their economics class appears to have allowed the economics students to effectively bridge the gap between them and the science students in this particular lexical domain. These findings support both the relationship between language level and LA, and also the positive effect that CLIL has on students' content-related vocabulary (Dalton-Puffer, 2008; Heras & Lasagabaster, 2015). The longitudinal

analysis indicated that there was a significant main effect of CLIL group on the English C-test at each data collection, with science students performing better than the economics group overall, and a significant main effect of time, with participants improving from 10th grade to 11th grade. There was, however, no significant interaction between time and CLIL group, as both groups increased their score from 10th grade to 11th grade. In the case of the English LAT, there was again a significant main effect of CLIL group, with the science group producing more tokens than the economics group overall. However, there was no significant main effect of time, and no significant interaction between time and CLIL group. Thus, while the science group produced a higher number of tokens in each grade, means at each data collection were relatively consistent, with minor improvements which were unrelated to CLIL group. Similar results were found in the case of the individual prompts, with a significant main effect of CLIL group in the first four prompts at each time, and science students producing more tokens than economics students in all cases. However, no significant main effect of CLIL group was observed for the prompt *Economy & Money*. Thus, although the science group performed better from 10th to 11th grade than economics students in all other cases, in the case of this content-relevant prompt both groups performed similarly, which is again likely attributable to the economics students' exposure to content-related vocabulary in their CLIL class. As was the case for the overall LAT, there was no significant main effect of time and no significant interaction between time and CLIL group for any of the five prompts of the English LAT, indicating that means at each data collection were relatively similar, with minor improvements which were unrelated to CLIL group. The longitudinal results thus again point to the value of CLIL teaching, given that no significant main effect of CLIL group was observed for the prompt *Economy & Money*, despite the fact there a significant main effect of CLIL group on the English C-test, the English LAT, and the other four prompts of the LAT, with science students outperforming economics students in all cases. These findings are consistent with recent research which highlights the benefits of CLIL in terms of the acquisition of subject-specific vocabulary. For example, Rieder-Bünemann, Hüttner and Smit (2022) analysed the spontaneous oral classroom productions of secondary school students in Austria taking European economics and politics through English. Their results show that there was substantial and active use of subject-specific terminology, though there was considerable variation between different students.

In terms of motivation, there were no statistically significant difference between

the two groups in 10th grade, and differences were found only in the categories *The “Ought to” Self* and *L2 Self Confidence* in 11th grade: the economics students reported higher motivation towards external sources but had lower self-confidence in English than the science students. Given that the economics students in 11th grade continued to study this subject through English, whereas the science students did not, it is very logical that economics students may feel more pressure from external sources, for example, teachers and parents, to do well in English. In addition, their lower self-confidence may well be attributed to their lower performance and language level, which is likely something that they are aware of based on their school grades. This has recently been highlighted as an issue by Sangeetha and Mekala (2021), who note that grades impact self-image and self-confidence, affecting expectations and influence, and particularly so in the case of older, high school students. Despite these minor differences, it appears that in this study, the linguistic advantage of the science students may not be attributed to higher motivation, as was hypothesised. In addition, regarding the longitudinal analysis, no statistically significant differences were found between the groups in either grade, indicating that the differences between science and economics students in language level and LA may not be attributed to language learning motivation. It is thus suggested that in this case, CLIL plays a greater role than motivation in influencing the results of the content-related prompt *Economy & Money*.

9.4.2. Qualitative Differences in CLIL: Lexical Availability

With regard to the qualitative differences in the words produced in English of students taking different CLIL classes at each testing period, results revealed very similar results for the science and economics students in 10th and 11th grade.

Regarding the frequency of first word responses, only one prompt differed between the groups, namely *Economy & Money* in 10th grade, where science students' most frequent first word was “economy” and economics students' most frequent first word was “money”, notably both words which feature in the title of the prompt itself. However, it was also noted that there was more variety in the first word produced for this prompt for the economics group: aside from the three students who produced “money”, other frequent first words included “risk” and “income”, which are clearly more representative of the type of target vocabulary one would expect students to be exposed to in an economics class.

Regarding the most and least productive prompts, results were very similar for

both groups: *Environment and Climate* was usually the most productive content-relevant prompt, followed by *Sport & Physical Activities* and then *Economy & Money*. However, economics students in 10th grade produced more tokens for the prompt *Economy & Money* than for *Sport & Physical Activities*, perhaps due to the exposure to content-related vocabulary they received in their economics class. This was notably not found in 11th grade, where many participants no longer studied economics through the medium of English.

Regarding lexical sophistication based on the non-shared words, only minor differences were observed. In 10th grade, economics students produced slightly more non-shared words for the prompt *Environment & Climate* than science students. Another notable finding was that while the prompt *Economy & Money* contained the highest percentage of non-shared words for both groups in both grades, the percentage for economics students was slightly higher than that of science students in 11th grade, indicating a slightly higher lexical sophistication in this measure. Regarding lexical sophistication based on the number of infrequent words in the production of each prompt, greater differences were observed in 11th grade, where economics students produced fewer words at the K1-K5 level and more off-list words in the prompt *Economy & Money*, while science students were found to do so in the other two content-related prompts. This higher production of off-list words for this prompt again points to higher lexical sophistication on the part of the economics students. These findings are consistent with recent research comparing CLIL and non-CLIL students' writings. For example, Lee (2020) analysed the written language competence of 11th grade CLIL and non-CLIL Korean students. Results revealed that those in the CLIL group demonstrated higher lexical sophistication, as well as mean length of clause, lexical diversity and writing quality. Similarly, in a Spanish context, Lahuerta Martínez (2020) analysed fluency, accuracy, grammatical complexity and lexical complexity in the writings of 7th and 10th grade CLIL and non-CLIL students. CLIL students again outperformed non-CLIL students overall and on all aspects measured. In the current study, however, what is all the more interesting is that this CLIL advantage appears to be topic-related, in that the students improve their lexical sophistication in the specific lexical domain which they are exposed to in their CLIL classes.

Drawing together these qualitative results, while it is clear that the students generally perform rather similarly regardless of CLIL group, there are some notable differences across the four measures of the qualitative analysis in the prompt *Economy &*

Money, which is notably where it was expected that differences would arise. This included, for the economics students, more variety in the first word produced, more tokens for the prompt *Economy & Money* than for *Sport & Physical Activities* in 10th grade, a higher percentage of non-shared words in 11th grade, and fewer words at the K1-K5 level and more off-list words. These findings suggest that, despite the linguistic and lexical advantages on the part of the science students, discussed in the previous section, economics students do appear to produce content-related vocabulary which is qualitatively different to that of the science students. This again supports suggestions that CLIL instruction has a positive effect on students' content-related vocabulary (Dalton-Puffer, 2008; Heras & Lasagabaster, 2015). While in this case the participants at hand enrol in economics in English only for one to two years, it would be extremely interesting to observe differences between elective CLIL subjects over a longer period of time, so as to better determine the effect of this exposure in the long run.

9.4.3. Quantitative Differences in CLIL: Content-Relevant Prompts

With regard to the quantitative differences in the words produced by students in English and French when the prompt was related to a CLIL class taken in either English (Physical Education and/or Biology/Physics and Chemistry) or French (Geography and History) at each testing period, results indicated that the participants retrieved a higher number of tokens in the English LAT, both overall and in all five prompts, irrespective of whether the prompt was related to an English or French CLIL class. Given suggestions by Schmitt (2000) that there is a relationship between language proficiency and the number and type of words retrieved in word association tasks, it is assumed these findings can be attributed to the students' language level, which was much higher in English than in French. Potential advantages that may come from studying geography and history in French are thus not observed. It is also possible that the prompt chosen to tap into the content-related vocabulary of the students' French CLIL class did not lead to the expected production. It is thus suggested that it would be beneficial for future research first and foremost to attempt to compare students with comparable language proficiencies in each TL. In addition, it would be valuable when choosing prompts to attempt to address specific units which have been studied in the CLIL classes, in other words, to home in on the specific target vocabulary which students are expected to have acquired in their lessons.

Despite the above findings, the longitudinal analysis does, however, point to

interesting advantages. Specifically, although there were no statistically significant differences in any of the content-relevant prompts in the English LAT from 10th to 11th grade, statistically significant differences were found in the one French content-relevant prompt *Environment & Climate*. Thus, despite the fact that studying geography through French does not result in the production of more tokens in the prompt *Environment & Climate* in French as compared to English, it does appear that there is a benefit to the students' exposure to content-related vocabulary, which becomes clear in the longitudinal analysis. This again highlights the need for research to carry out this comparison over a longer period, in order to determine how exposure to content-related vocabulary may improve across time.

9.4.4. Qualitative Differences in CLIL: Content-Relevant Prompts

With regard to the qualitative differences in the words produced by students in English and French when the prompt was related to a CLIL class taken in either English or French at each testing period, results indicated remarkable similarities across the two languages, regardless of the content-relevant prompt at hand.

Regarding the most frequent first word for each prompt, few differences were observed across the content-relevant prompts. For the prompt *Sport & Physical Activities*, “football” was the most frequent first word in both languages at all levels, with the exception of 10th grade students in French, where “basketball” was the most common. For the prompt *Environment & Climate*, “sun” was the most common word in both languages at all levels. For the prompt *Economy & Money*, greater differences were observed, though in all cases the most common first word was either a currency (“dollar” in English and “euro” in French for 9th grade students) or found in the title of the prompt itself (“economy” in both languages 10th grade students and “money” in both languages for 11th grade students). Notably, however, regarding the difference between the two TLs, practically no differences were observed, with the exception of “football” / “basketball” in *Sport & Physical Activities* in 10th grade and “dollar” / “euro” in *Economy & Money* in 9th grade.

Regarding the ranking of the most and least productive content-relevant prompts, the ranking was precisely the same in both languages for 9th and 11th grade students and in English for 10th grade students, where the highest number of words were retrieved in the prompt *Environment & Climate*, followed by *Sport & Physical Activities* and then *Economy & Money*. However, in French in 10th grade, *Economy & Money* was again the

prompt that produced the fewest words, but *Sport & Physical Activities* was found to be the most productive, followed by *Environment & Climate*. Thus, though 10th grade students studied geography through French and physical education through English, *Environment & Climate* was more productive in English and *Sport & Physical Activities* was more productive in French, that is, the reverse of what would be expected. However, it should be again noted that for the prompt *Sport & Physical Activities* in French, 76% of the words produced were English cognates, which may likely account for the higher number of words produced in this prompt. This has also been found in recent research into receptive knowledge of L3 French of Swedish learners by Lindquist (2020): when analysing the receptive vocabulary of students from 6th to 9th grade, results revealed that a significant proportion of known words were cognates, either in the students' L1 (Swedish), L2 (English), or both. Thus, it appears that in the case of their L3, students may always rely on the vocabulary they have been exposed to in their English CLIL classes.

Regarding non-shared words, clear similarities were again found across both languages, with slight differences in 10th grade. Percentages of non-shared words in 9th and 11th grade were the same in both languages: in 9th grade *Economy & Money* contained the highest percentage, followed by *Sport & Physical Activities* and then *Environment & Climate*, while in 11th grade *Economy & Money* contained the highest percentage, followed by *Environment & Climate* and then *Sport & Physical Activities*. In 10th grade, *Economy & Money* again had the highest number of non-shared words in both languages; however, in English this was followed by *Sport & Physical Activities* and then *Environment & Climate*, while in French it was followed by *Environment & Climate* and then *Sport & Physical Activities*. Notably, in this case, students retrieved a higher number of non-shared words in *Sport & Physical Activities* in English, which was related to physical education studied through English, and a higher number of non-shared words in *Environment & Climate* in French, which related to geography studied through French. Aside from this difference, however, results generally point to no difference in non-shared words dependant on the language of instruction.

Regarding the number of infrequent words produced in each language, despite the differences in the corpora used, which should evidently be taken into consideration, clear parallels were found in both languages across all grades. In both English and French, the highest percentage of words in the K-1 to K-5 lists were found in the prompt *Environment & Climate*, followed by *Economy & Money* and then *Sport & Physical Activities*. In

addition, the prompt *Sport & Physical Activities* contained the highest number of off-list words, which as previously mentioned may be attributed to the appearance of multiword units in this lexical domain. Thus, there again seems to be no clear difference between the participants' lexical sophistication which may be attributed to CLIL instruction.

In sum, the qualitative analysis of the tokens produced by participants in English and French when the prompt was related to a CLIL class taken in either English or French at each testing period revealed no clear differences. One exception was that in 10th grade, a higher number of non-shared words were found in *Sport & Physical Activities* than *Environment & Climate*, while in French a higher number of non-shared words were found in *Environment & Climate* than *Sport & Physical Activities*. However, given that this was the only indication of a difference between the two languages related to the participants' CLIL classes, it can be concluded that the language of instruction made no clear qualitative impact on the students' content-relevant prompts in English and French on the basis of the LATs.

Chapter 10: Summary and Conclusions

This final chapter will open by providing the main conclusions of the research, highlighting the key research findings and the contributions that they offer to the field. A translation of these conclusion into Spanish is provided in Appendix I. It will then discuss the limitations of the research that has been carried out, indicating some key issues which should be taken into account when interpreting the findings. It will close by suggesting some directions for further research.

10.1. Conclusions

This doctoral thesis has aimed to meet a number of objectives in relation to LA, language learning motivation, gender and CLIL in two TLs, English and French. In particular, it attempted to investigate:

- the quantitative and qualitative differences between the participants' LA in English and French
- the quantitative differences between the participants' English language learning motivation as compared with their French language learning motivation
- quantitative and qualitative differences between male and female participants' LA in English and French and quantitative differences between their language learning motivation in each language
- the impact of CLIL instruction on the participants' LA and language learning motivation in each language

In order to do this, it analysed the LA, language learning motivation and language level of 9th, 10th and 11th grade students enrolled simultaneously in CEIL and CLIL both in English and French by means of LATs, language learning motivation questionnaires and C-tests.

The first research question allowed us to conclude that there are clear differences between the participants' LA in English and French, to the advantage of English, which may likely be due to the participants' proficiency in each language. The cross-sectional analysis also revealed the advantage of CLIL in terms of exposing students to content-related vocabulary, for example, 10th grade students, who had begun to study economics in English, produced a higher number of tokens in the prompt *Economy and Money* in

both languages than 9th grade students. Meanwhile, the longitudinal analysis revealed that there was a greater difference in the participants' LA in French than in English from 10th to 11th grade, perhaps due to the participants' lower level in French. The qualitative analysis, on the other hand, allowed us to conclude that there was a large degree of similarity in each language. However, some clear differences were found in the ranking of the most and least productive prompts, which appear to be dependent on the language at hand. While English content-relevant prompts were the least productive at all three grades in the English LAT, the French content-relevant prompt was found to be the most productive for 9th grade and 11th grade students in the French LAT. Thus, exposure in French CLIL appears to have made more of a difference than English. These findings add to the field of LA firstly by answering calls by Canga Alonso (2017) and highlighting that there is indeed a relationship between language level and LA, and secondly by showing the variation between different TLs such as English and French, an issue which previous research has scarcely addressed.

Regarding the second research question, we could conclude that there is a clear difference between students' motivation towards English and French, with participants' indicating a higher level of motivation towards English than to French in all cases. This difference remained constant across the different grade levels, with few differences between students in 9th and 10th grade and between students in 10th and 11th grade. However, in the latter, we could also see a more negative attitude towards learning in English in 11th grade, which is potentially related to the general removal of CLIL teaching in this grade. Analysis of the relationship between LA, motivation and language level also allowed us to observe clear differences in each language, as there appears to be a much stronger relationship between LA and motivation and between language level and motivation in English than in French. These results contribute to the investigation of motivation towards EFL and LOTEs, answering Oakes and Howard's (2019) call for both motivation research on LOTEs and better integration between research on LOTEs and EFL, highlighting the key differences in the two TLs.

The third research question allowed us to infer that across languages, participants produced a higher number of words in English than in French regardless of gender, while within languages, female students produced a higher number of words than male students in the overall LAT in almost all cases. We can thus see that the English advantage remains for both male and female students, while female students generally perform better than male students in both languages. In terms of the individual prompts, we could also

observe that CLIL instruction appear to play a part, for example, no differences were found in the prompt *Economy & Money* in 11th grade, when male students continued to study economics in English, while female students did not. We could also see that while both male and female students improved from one grade to the next, they did so in different prompts and languages. Qualitatively speaking, few differences arose between male and female students' LA in English and French. However, results did suggest that while male students produce fewer words than female students, they may demonstrate greater lexical sophistication based on their higher percentage of non-shared words and higher percentage of off-list words. These findings build on previous research into gender and LA, offering in particular interesting observations concerning the different ways that male and female students progress across the three grades at hand, the effect of CLIL, and the potential quantitative and qualitative differences. In terms of motivation, across languages we can again conclude that participants are more motivated towards English regardless of gender, while within languages, male and female students were equally motivated towards learning English, but clearer differences arose in French, particularly in 10th grade. We also discovered indications that male students very often report higher motivation towards CLIL subjects than female learners in both languages. These findings make an important contribution to gender studies, motivation and CLIL, as they support the suggestion by Heras and Lasagabaster (2015) that the learning context affects male and female students' language learning motivation in a different way.

Finally, the fourth research question provided interesting insights into the effect of CLIL instruction, specifically regarding the subject economics which was related to the LA prompt *Economy & Money*. Of the utmost importance is the fact that CLIL instruction appears to allow students to improve their LA in this prompt to the extent that they can effectively bridge the gap between them and other students who do not study this content-related vocabulary in the TL. This is the case even when other students have a higher language level and perform better in other lexical domains. In addition, motivation was not found to play a part in these differences. Qualitatively speaking, CLIL instruction also appears to result in more variety in the first word produced, as well as the production of fewer words at the K1-K5 level and more off-list words in the prompt *Economy & Money*, indicating higher lexical sophistication. Regarding the comparison of LA in English and French when a prompt is related to a CLIL class taken in either English or French, we could conclude that participants produced a higher number of words in the English LAT, regardless of whether the prompt at hand was related to an

English or French CLIL class. In this case, this was likely due to the large difference in language proficiency, given that the participants' French language level was considerably lower than their English language level. However, there is some indication that CLIL language does have an effect on LA: no differences were found between 10th and 11th grade students in English, yet there was a difference in the one French content-relevant prompt *Environment & Climate*. Thus, although the difference in language levels may conceal any advantage, it does appear that the language of instruction helps to improve LA in that language in related prompts. Qualitatively speaking, however, the language appears to have had no major impact on the students' content-relevant prompts in English and French. The above results are incredibly important for the field of CLIL, as they answer Canga Alonso's (2017) call for a focus on LA prompts which may be relevant to the CLIL subject, demonstrating the advantage of this learning context in terms of content-related vocabulary. They also show the important effect that language level can have, which should evidently be taken into consideration when comparing CEIL and CLIL, answering calls for the comparison of different TLs in this context (Dalton-Puffer, Nikula & Smit, 2010; Cenoz et al, 2014; Merino & Lasagabaster, 2018a).

The above findings are of value to both researchers and teachers alike. From a research perspective, the study has answered several calls that have been made in previous research and has filled a number of gaps in the research which have been highlighted, further expanding our understanding of these intricate issues. In addition, educators can benefit from these findings, which highlight key considerations in the context of multilingual CLIL instruction. In this context where a lack of teacher preparation has been continuously flagged, the results of this study can provide CLIL stakeholders with the necessary information to better prepare and provide a more beneficial teaching context.

10.2. Limitations

There are a number of limitations which need to be taken into consideration when interpreting the findings of this research.

Firstly, the study had a relatively small sample size of 91 participants. In addition, many of the subgroups which were analysed were rather small, for example, there were between 40 to 42 students in each grade, and so when analysing gender within each grade, groups necessarily included fewer participants. While requirements for statistical consideration were met, it would evidently be valuable to include a larger sample size in order to strengthen the results.

Another matter related to the sample which should be born in mind is the socio-economic status (SES) of the participants involved. This is because SES has been found to be related to L2 achievement, with learners with higher SES outperforming those with lower SES (Ellis, 2008). Though much research into LA which has been carried out in the region of La Rioja has been in public schools, in this study it was done in semi-private schools. This was necessary given the specific requirements of the research at hand, that is, to investigate the differences between students' simultaneously studying English and French in a CLIL context, as the chosen schools were the only ones in the area which had this profile. While all participants will likely have had a similar SES, this characteristic should be considered if we are to compare the findings with other similar research in the region.

A further issue to bear in mind is the fact that while the first data collection took place under relatively normal circumstances, the second data collection took place amid the COVID-19 pandemic. While the students had returned to normal classes one semester prior to the second data collection, it is of course likely that students in 9th and 11th grade experienced distributions to their normal school day, and in addition that the ongoing situation may have affected them and their performance in the tests. This should be taken into consideration when comparing these students with those in 10th grade, as the conditions were necessarily different at each point in time.

Another clear limitation of the study which has been previously addressed is the participants' language level in English and French. Although the schools' policy states clearly that students' school day is divided up in such a way that they receive a third of their instruction in Spanish, a third in English and a third in French, in practice this did not seem to be the case. Instead, there was a clear focus on English, in which students received instruction in up to five different subjects, as compared to French, in which students received instruction in only two subjects. There is thus a discrepancy in the number of hours of exposure that students receive in each of their TL. In order to adequately compare two TLs, it is clearly necessary to ensure that students receive equal exposure to each one, in order to rule out the likelihood that differences which arise are simply attributable to hours of exposure.

Another limitation concerns the research methodology. As is highlighted by Wei and Moyer (2009), some issues with a cross-sectional design in the study of multilingual learners is that it is not suitable for tracing a sequential developmental pattern of particular subjects or change over time, it lacks the ability to assess individual differences, and there

is weaker evidence of causality. While inclusion of the 9th grade group in this study made it possible to investigate LA, motivation and gender across three school grades instead of two, it would evidently have been more beneficial to be able to carry out a longitudinal study across the three grades. This would have made it possible to determine the sequential development of the same learners and to rule out the possibility that differences which arose could be attributable to individual differences in the groups under analysis.

With regard to the LAT, this research attempted to take some preliminary steps towards catering LA prompts to the CLIL subjects in which students are enrolled. However, it is suggested that some prompts, such as *Environment & Climate*, may not have been optimal in triggering the content-related vocabulary covered in the students' CLIL class (Geography and History). It may be more advantageous to begin with a more thorough examination of the specific content which students are to cover, for example, by analysing the textbooks that are used in their CLIL classes, so as to better tap into specific areas of their CLIL classes via the LA prompts.

In terms of language learning motivation, a clear disadvantage of the methodology used is that it took only a quantitative approach. Although it was beyond the scope of this study, it would have been extremely beneficial to adopt a qualitative approach, so as to better understand the shifting motivations of the students across the three grade levels.

A final issue which should be mentioned is the differences in the CLIL subjects across the three grades under analysis, and at times for male and female students or different groups of students. While one advantage of this is that it allowed for the comparison of different lexical exposure in different stages and in different groups of students, one drawback is that it prevented us from analysing the long-term effect of CLIL across numerous grade levels. For example, in the case of the subject economics: 9th grade students did not study this subject, a subgroup of 10th grade students studied this subject in English, and a subgroup of 11th grade male students studied this subject in English while a subgroup of female students studied it in Spanish. While this allowed us to compare students who received no exposure to students who received exposure, it would also have been beneficial to be able to follow a larger cohort of students studying the subject in the TL for a longer period of time, in order to better determine the long-term effects of CLIL exposure.

10.3. Directions for Further Research

Considering the key findings of this research and the main limitations discussed above, a number of suggestions can be made for future research.

Firstly, regarding the comparison of English and French, it would first and foremost be valuable for future research to compare students with comparable language proficiencies and hours of exposure in each TL, as discussed above. This would allow us to determine whether the differences observed are truly attributable to the languages themselves, rather than language level. It would also be interesting to investigate other languages alongside English, including minority and less-spoken languages. When comparing different FLs, it is also suggested that future research into motivation home in on the compulsory nature of learning the TLs at hand, as this is very likely an important factor that needs to be taken into consideration.

Regarding LA and gender, one interesting finding which needs further attention is the fact that while female participants in this study outperformed their male peers in the LAT, the results also suggest that male participants produce words which are more lexically sophisticated. Future research should aim to clarify this finding and determine whether the female advantage can be attributed to the use of more simple vocabulary.

Concerning language learning motivation, it is firstly suggested that future research endeavour to incorporate a qualitative approach alongside a quantitative analysis, for example, conducting interviews with a small subgroup of students on the basis of the findings of the MFQ. This would allow us to paint a more complete picture of the learners' motivational profiles, benefiting from the advantages which each methodological approach has to offer. In addition, given findings which point to an increase in anxiety towards learning English in older adolescent learners, future research would benefit from investigating the reasons behind these differences throughout secondary and high school. This would allow us to confirm whether age is an important factor in this regard.

Regarding CLIL, future research into LA and motivation would benefit from adopting more specifically catered instruments in the analysis. In the former, efforts should be made to link LA prompts to the specific units which have been studied in the CLIL classes, in order to target the specific target vocabulary in the participants' curriculum. In the latter, when investigating attitudes towards learning, it is recommended that further questions should be incorporated addressing each CLIL subject, so as to increase the internal reliability of the scales and better understand the effect of learners' attitudes toward their CLIL classes.

Finally, as has been previously mentioned, it would be extremely beneficial for future research to conduct further longitudinal research across a longer period of time, tracking the same cohort of students enrolled in the same CLIL subjects over several years. This would provide us with a more detailed picture of the progress of the participants as well as the effect that CLIL has on their linguistic development.

References

- Ackerl, C. (2007). Lexico-grammar in the essays of CLIL and non-CLIL students: Error analysis of written production. *Vienna English Working Papers (Views) (Special Issue: Current Research on CLIL 2)*, 3(16), 6-12.
- Agustín Llach, M. P. (2009). The role of Spanish L1 in the vocabulary use of CLIL and non-CLIL EFL learners. In R. M. Jiménez Catalán & Y. Ruiz de Zarobe (Eds.), *Content and language integrated learning: Evidence from research in Europe* (pp. 112-129). Bristol: Multilingual Matters.
- Agustín Llach, M. P. (2010). Exploring the role of gender in lexical creations. In R. M. Jiménez Catalán (Ed.), *Gender Perspectives on Vocabulary in Foreign and Second languages* (pp. 74-92). New York: Palgrave Macmillan.
- Agustín Llach, M. P. (2014). Exploring the lexical profile of young CLIL learners: Towards an improvement in lexical use. *Journal of Immersion and Content-Based Language Education*, 2(1), 53-73.
- Agustín Llach, M. P. (2016). Age and type of instruction (CLIC vs. traditional EFL) in lexical development. *International Journal of English Studies*, 16(1), 75-96.
- Agustín Llach, M. P. (2017). The effects of the CLIL approach in young foreign language learners' lexical profiles. *International journal of bilingual education and bilingualism*, 20(5), 557-573.
- Agustín Llach, M. P., & Canga Alonso, A. (2016). Vocabulary growth in young CLIL and traditional EFL learners: evidence from research and implications for education. *International Journal of Applied Linguistics*, 26(2), 211-227.
- Agustín Llach, M. P., & Fernández Fontecha, A. (2014). Lexical variation in learners' responses to cue words: The effect of gender. In R. M. Jiménez Catalán (Ed.), *Lexical availability in English and Spanish as a second language* (pp. 69-81). Dordrecht: Springer.
- Agustín Llach, M. P., & Jiménez Catalán & R. M. (2007). Lexical reiteration in EFL young learners' essays: does it relate to the type of instruction? *International Journal of English Studies*, 7(2), 85-104.

- Agustín Llach, M. P., & Terrazas Gallego, M. (2012). Vocabulary knowledge development and gender differences in a second language. *Elia*, 12(1), 45-75.
- Agustín-Llach, M. P. (2019). The impact of bilingualism on the acquisition of an additional language: Evidence from lexical knowledge, lexical fluency, and (lexical) cross-linguistic influence. *International Journal of Bilingualism*, 23(5), 888-900.
- Agustín-Llach, M. P. (2022). How age and L2 proficiency affect the L2 lexicon. *System*, 104, 102697.
- Aiken, L. (1997). *Questionnaires and inventories: Surveying opinions and assessing personality*. New York: John Wiley.
- Alderman, M. K. (2013). *Motivation for achievement: Possibilities for teaching and learning*. New York: Routledge.
- Alderson, J. C. (1979). The cloze procedure and proficiency in English as a foreign language. *TESOL Quarterly*, 13(2), 219-227. (Reprinted 1983 in J. W. Oller (Ed.) *Issues in Language Testing Research* (pp. 205-212), Rowley, Mass: Newbury House).
- Alejo González, R., & Piquer Píriz, A. M. (2016a). Measuring the productive vocabulary of secondary school CLIL students: Is Lex30 a valid test for low-level school learners? *Vigo International Journal of Applied Linguistics*, 13, 31-54.
- Alejo González, R., & Piquer-Píriz, A. (2016b). Urban vs. rural CLIL: An analysis of input-related variables, motivation and language attainment. *Language, Culture and Curriculum*, 29(3), 245-262.
- Al-Hoorie, A. H. (2017). Sixty years of language motivation research: Looking back and looking forward. *SAGE Open*, 7(1), 1-11.
- Al-Hoorie, A. H. (2018). The L2 motivational self system: A meta-analysis. *Studies in Second Language Learning and Teaching*, 8(4), 721-754.
- Allen, H. W. (2010). Language-learning motivation during short-term study abroad: An activity theory perspective. *Foreign Language Annals*, 43(1), 27-49.
- Alrabai, F. (2016). The effects of teachers' in-class motivational intervention on learners' EFL achievement. *Applied Linguistics*, 37(3), 307-333.

- Ander Merino, J., & Lasagabaster, D. (2018a). CLIL as a way to multilingualism. *International Journal of Bilingual Education and Bilingualism*, 21(1), 79-92.
- Ander Merino, J., & Lasagabaster, D. (2018b). The effect of content and language integrated learning programmes' intensity on English proficiency: A longitudinal study. *International Journal of Applied Linguistics*, 28(1), 18-30.
- Anderson, C. E. (2011). CLIL for CALP in the multilingual, pluricultural, globalized knowledge society: Experiences and backgrounds to L2 English usage among Latin American L1 Spanish-users. *Latin American Journal of Content & Language Integrated Learning*, 4(2), 51-66.
- Arribas, M. (2016). Analysing a whole CLIL school: Students' attitudes, motivation, and receptive vocabulary outcomes. *Latin American Journal of Content & Language Integrated Learning*, 9(2), 267-292.
- Artigal, J. M. (1993). Catalan and Basque immersion programmes. In H. Baetens Beardsmore (Eds.), *European models of bilingual education* (pp. 30-53). Clevedon: Multilingual Matters.
- Aslan, O. (2009). *The role of gender and language learning strategies in learning English* (Unpublished master's Thesis). Ankara: Middle East Technical University.
- Attard-Montalto, S., & Walter, L. (2021). *The CLIL4U Guidebook v.2 2021 with Technology Enhanced Learning*. Retrieved from <https://languages.dk/archive/clil4u/book/FINAL%20-%20THE%20UPDATED%20CLILSTORE%20GUIDEBOOK%20-%20DECEMBER%202021.pdf>
- Aviad-Levitzky, T., Laufer, B., & Goldstein, Z. (2019). The New Computer Adaptive Test of Size and Strength (CATSS): Development and Validation. *Language Assessment Quarterly*, 16(3), 345-368.
- Bacon, S. M. (1992). The relationship between gender, comprehension, processing strategies, and cognitive and affective response in foreign language listening. *The Modern Language Journal*, 76(2), 160-178.
- Badstübner, T., & Ecke, P. (2009). Student expectations, motivations, target language use, and perceived learning progress in a summer study abroad program in Germany. *Die Unterrichtspraxis/Teaching German*, 42(1), 41-49.

- Baetens Beardsmor, H. (2001). The Past Decade and the Next Millennium. In D. Marsh, A. Maljers & A. K. Hartiala (Eds.), *Profiling European CLIL Classrooms* (pp. 10-11). Jyväskylä, Finland: Centre for Applied Language Studies.
- Bailey, K. (1986, Spring). *Class lecture*. Monterey Institute of International Studies.
- Baker, S. C., & MacIntyre, P. D. (2000). The role of gender and immersion in communication and second language orientations. *Language Learning*, 50(2), 311-341.
- Ball, P. (2009). Does CLIL work? In D. Hill and A. Pulverness (Eds.), *The Best of Both Worlds?: International Perspectives on CLIL* (pp. 32-43). Norwich: Norwich Institute for Language Education.
- Ball, P., Kelly, K., & Clegg, J. (2016). *Putting CLIL into Practice: Oxford Handbooks for Language Teachers*. Oxford: Oxford University Press.
- Barbu, S., Nardy, A., Chevrot, J. P., Guellai, B., Glas, L., Juhel, J., & Lemasson, A. (2015). Sex differences in language across early childhood: Family socioeconomic status does not impact boys and girls equally. *Frontiers in Psychology*, 6, 1874,1-10.
- Barcroft, J. (2004). Second language vocabulary acquisition: A lexical input processing approach. *Foreign Language Annals*, 37(2), 200-208.
- Barcroft, J. (2007). When knowing grammar depends on knowing vocabulary: Native-speaker grammaticality judgements of sentences with real and unreal words. *Canadian Modern Language Review*, 63(3), 313-343.
- Barcroft, J. (2012). *Input-Based Incremental Vocabulary Instruction*. Alexandria, VA: TESOL International Association.
- Bardel, C., & Lindqvist, C. (2011). Developing a lexical profiler for spoken French L2 and Italian L2: The role of frequency, thematic vocabulary and cognates. *EUROSLA Yearbook*, 11(1), 75-93.
- Bardel, C., Gudmundson, A., & Lindqvist, C. (2012). Aspects of Lexical Sophistication in Advanced Learners' Oral Production: Vocabulary Acquisition and Use in L2 French and Italian. *Studies in Second Language Acquisition*, 34(2), 269-290.

- Barrios, E., & Milla Lara, M. D. (2020). CLIL methodology, materials and resources, and assessment in a monolingual context: an analysis of stakeholders' perceptions in Andalusia. *The Language Learning Journal*, 48(1), 60-80.
- Baten, K., van Hiel, S., & de Cuypere, L. (2020). Vocabulary development in a CLIL context: A comparison between French and English L2. *Studies in Second Language Learning and Teaching*, 10(2), 307-336.
- Bauer, L., & Nation, P. (1993). Word families. *International journal of Lexicography*, 6(4), 253-279.
- Beglar, A., & Hunt, A. (1999). Revising and validating the 2000 word level and the university word level vocabulary tests. *Language Testing*, 16(2), 131–162.
- Beglar, D. (2010). A Rasch-based validation of the vocabulary size test. *Language Testing*, 27(1), 101–118.
- Bell, C. M., McCarthy, P. M., & McNamara, D. S. (2006). Variations in language use across gender: Biological versus sociological theories. In R. Sun & N. Miyake (Eds.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 1009–1013). Austin, TX: Cognitive Science Society.
- Berman, I. M., Buchbinder, V. A., & Beznedezych, M. L. (1968). Formirovanie potencialnogo slovarnogo pri obučenii russkomu jazyku kak inostannomu [Forming a potential vocabulary in teaching Russian as a foreign language]. *Russkij jazyk za rubezhom*, 4, 57– 60.
- Bernaus, M., Wilson, A., & Gardner, R. C. (2009). Teachers' motivation, classroom strategy use, students' motivation and second language achievement. *Porta Linguarum: Revista Internacional de Didáctica de las Lenguas Extranjeras*, 12, 25-36.
- Berns, M., de Bot, K., & Hasenbrink, U. (2006). *English, media and youth in Europe*. New York: Kluwer.
- Bialystok, E., & Sharwood Smith, M. (1985). Interlanguage is not a state of mind: An evaluation of the construct for second-language acquisition. *Applied Linguistics*, 6(2), 101-117.
- Block, D., & Cameron, D. (2002). Introduction. In Da. Block & D. Cameron (eds.), *Globalization and language teaching* (pp. 1-10). London: Routledge.

- Bloomer, A. (2010). Designing a questionnaire. In S. Hunston & D. Oakley (Eds.), *Introducing Applied Linguistics Concepts and Skills* (pp. 145–150). London: Routledge.
- BOE (2020). *Ley Orgánica 3/2020, de 29 de diciembre, por la que se modifica la Ley Orgánica 2/2006, de 3 de mayo, de Educación*. Boletín Oficial del Estado No. 340, 30 December.
- Boo, Z., Dörnyei, Z., & Ryan, S. (2015). L2 motivation research 2005–2014: Understanding a publication surge and a changing landscape. *System*, 55, 145-157.
- BOR (2018). *Orden EDU/31/2018, de 1 de junio, por la que se regulan los centros de Educación Secundaria, Bachillerato y Formación Profesional bilingües en la Comunidad Autónoma de La Rioja*. Boletín Oficial de la Rioja No. 67, 8 June.
- Bowden, H. W., Sanz, C., & Stafford, C. A. (2005). Individual differences: Age, sex, working memory, and prior knowledge. In C. Sanz (Ed.), *Mind and context in adult second language acquisition: Methods, theory, and practice* (pp. 105-140). Washington, DC: Georgetown University Press.
- Brady, I. K. (2014). L2 motivation in Spanish university students: A pilot study on the L2 motivation self system. In G. Alcaráz-Mármol, & M. M Jimenez-Cervantes (Eds.), *Studies in Philology: Linguistics, literature and cultural studies in modern languages* (pp. 95-116). Newcastle upon Tyne: Cambridge Scholars Press.
- Brady, I. K. (2015). *The ideal and ought-to L2 selves of Spanish learners of English*. (Unpublished doctoral dissertation). Universidad de Murcia, Spain.
- Brady, I. K. (2019). Exploring the L2 Motivational Self System in Spain: Study design and preliminary findings. *Journal of English Studies*, 17, 43-73.
- Breidbach, S., & Viebrock, B. (2012). CLIL in Germany: Results from recent research in a contested field of education. *International CLIL Research Journal*, 1(4), 5-16.
- Breiteneder, A. (2005). *Exploiting redundancy in English as a European lingua franca: The case of the 'third person -s'*. (Unpublished master's Thesis). Vienna: University of Vienna.
- Broca, Á. (2016). CLIL and non-CLIL: Differences from the outset. *ELT Journal*, 70(3), 320-331.

- Brown, D. (2013). Types of words identified as unknown by L2 learners when reading. *System*, 41(4), 1043-1055.
- Brown, D. (2018). Examining the Word Family through Word Lists. *Vocabulary Learning and Instruction*, 7(1), 14–34.
- Brown, H. D. (2007). *Principles of language learning and teaching* (5th ed.). New York: Pearson Longman.
- Brown, H., & Bradford, A. (2017). EMI, CLIL, & CBI: Differing approaches and goals. In P. Clements, A. Krause, & H. Brown (Eds.), *Transformation in Language Education* (pp. 328-334). Tokyo: JALT.
- Brown, J. D. (2004). Research Methods for Applied Linguistics: Scope, Characteristics, and Standards. In A. Davies & C. Elder (Eds.), *The handbook of applied linguistics* (pp. 476-500). Oxford: Blackwell.
- Bruton, A. (2011). Is CLIL so beneficial, or just selective? Re-evaluating some of the research. *System*, 39(4), 523-532.
- Bruton, A. (2012). Review of “Language use and language learning in CLIL classrooms” by C. Dalton-Puffer, T. Nikula & Ute Smit (Eds.). *Applied Linguistics*, 33(4), 456-459.
- Bruton, A. (2013). CLIL: Some of the reasons why... and why not. *System*, 41(3), 587-597.
- Bruton, A. (2015). CLIL: Detail matters in the whole picture. More than a reply to J. Hüttner and U. Smit (2014). *System*, 53, 119-128.
- Bui, G., & Teng, F. (2019). Exploring complexity in L2 and L3 motivational systems: a dynamic systems theory perspective. *The Language Learning Journal*, 1-16.
- Bulté, B., Surmont, J., & Martens, L. (2021). The impact of CLIL on the L2 French and L1 Dutch proficiency of Flemish secondary school pupils. *International Journal of Bilingual Education and Bilingualism*.
<https://doi.org/10.1080/13670050.2021.2018400>

- Burkett, T. (2015). An Investigation into the Use of Frequency Vocabulary Lists in University Intensive English Programs. *International Journal of Bilingual & Multilingual Teachers of English*, 3(02), 71-83.
- Busse, V. (2017). Plurilingualism in Europe: Exploring attitudes toward English and other European languages among adolescents in Bulgaria, Germany, the Netherlands, and Spain. *The Modern Language Journal*, 101(3), 566-582.
- Busse, V., & Walter, C. (2013). Foreign language learning motivation in higher education: A longitudinal study of motivational changes and their causes. *The Modern Language Journal*, 97(2), 435-456.
- Busse, V., & Williams, M. (2010). Why German? Motivation of students studying German at English universities. *Language Learning Journal*, 38(1), 67-85.
- Calafato, R., & Tang, F. (2019). Multilingualism and gender in the UAE: A look at the motivational selves of Emirati teenagers. *System*, 84, 133-144.
- Campillo, J. M., Sánchez, R., & Miralles, P. (2019). Primary Teachers' Perceptions of CLIL Implementation in Spain. *English Language Teaching*, 12(4), 149-156.
- Canga Alonso, A. (2013a). The receptive vocabulary of Spanish 6th-grade primary-school students in CLIL instruction: A preliminary study. *Latin American Journal of Content & Language Integrated Learning*, 6(2), 22-41.
- Canga Alonso, A. (2013b). The receptive vocabulary of Spanish students in CLIL and Non-CLIL instruction. *A Journal of Anglo-American Studies 2*, Second Series.
- Canga Alonso, A. (2015a). The receptive vocabulary size of Spanish 5th grade Primary School students in CLIL and Non-CLIL instruction. *ES Review. Spanish Journal of English Studies*, 36, 63-86.
- Canga Alonso, A. (2015b). Receptive vocabulary of CLIL and Non-CLIL primary and secondary school learners. *Complutense Journal of English Studies*, 23, 59-77.
- Canga Alonso, A. (2017). Spanish L1 speakers' and EFL learners' available lexicon. *Anuario de estudios filológicos*, 40, 5-23.
- Canga Alonso, A., & Arribas García, M. (2015). The benefits of CLIL instruction in Spanish students' productive vocabulary knowledge. *Encuentro: revista de investigación e innovación en la clase de idiomas*, 24, 15-31.

- Caraker, R. (2016). Spain and the context of English language education. *Research Bulletin*, 92, 23-35.
- Carroll, J. B. (1938). Diversity of vocabulary and the harmonic series law of word-frequency distribution. *Psychological Record*, 2, 379-386.
- Castellano-Risco, I. (2018). Receptive vocabulary and learning strategies in secondary school CLIL and non-CLIL learners. *Onomázein Revista de lingüística, filología y traducción de la Pontificia Universidad Católica de Chile*, 40, 28-48.
- Castellano-Risco, I., Alejo-González, R., & Piquer-Píriz, A. (2020). The development of receptive vocabulary in CLIL vs EFL: Is the learning context the main variable? *System*, 91, 102-263.
- Celaya, M. L., & Ruiz de Zarobe, Y. (2010). First languages and age in CLIL and non-CLIL contexts. *International CLIL Research Journal*, 1(3), 60-66.
- Cenoz, J. (2015). Content-based instruction and content and language integrated learning: The same or different? *Language, culture and curriculum*, 28(1), 8-24.
- Cenoz, J., Genesee, F., & Gorter, D. (2014). Critical analysis of CLIL: Taking stock and looking forward. *Applied Linguistics*, 35(3), 243-262.
- Chamberlain, A. (1965, November). *Leaning a Passive Vocabulary*. Paper presented at International Conference of Modern Foreign Language Teaching November. Part 1.
- Chambers, G. N. (1993). Taking the 'de' out of demotivation. *Language Learning Journal*, 7, 13-16.
- Chang, L. Y. H. (2010). Group processes and EFL learners' motivation: A study of group dynamics in EFL classrooms. *TESOL Quarterly*, 44(1), 129-154.
- Clément, R. (1980) Ethnicity, contact and communicative competence in a second language. In H. Giles, W.P. Robinson & P.M. Smith (Eds.), *Language: Social Psychological Perspectives* (pp.147-154). Oxford: Pergamon.
- Clément, R., Dörnyei, Z., & Noels, K. (1994). Motivation, self-confidence, and group cohesion in the foreign language classroom. *Language Learning*, 44, 417-448.

- Coates, L., & Johnson, T. (2001). Towards a social theory of gender. In W. Peter Robinson & Howard Giles (Eds.), *The New Handbook of Language and Social Psychology* (pp. 451-464). New York: Wiley.
- Cobb, T. (2018). Web vocabprofile v.4 [Computer program]. Retrieved from <https://www.lex tutor.ca/vp/eng/>
- Cobb, T. (2020). Range for Texts v.4 [Computer program]. Retrieved from <https://www.lex tutor.ca/cgi-bin/range/texts/>
- Cohen, A. D., Segal, M., & Weiss Bar-Siman-Tov, R. (1985). The C-test in Hebrew. In C. Klein-Braley & U. Raatz (Eds.), *C-Tests in der Praxis* (pp. 121-127). AKS-Rundbrief Special Issue 13/14. Bochum, Germany: Ruhr-Universität Bochum.
- Cohen, J. (1992). Statistical power analysis. *Current Directions in Psychological Science*, 1(3), 98-101.
- Consejería de Educación al Consejo Escolar (2010). *Los Programas de Enseñanza Bilingüe de la Comunidad de Madrid: Un estudio comparado*. Retrieved from <http://www.madrid.org/bvirtual/BVCM001973.pdf>
- Cook, V., & Singleton, D. (2014). *Key Topics in Second Language Acquisition*. Bristol: Multilingual Matters.
- Coonan, C. M. (2007). Insider views of the CLIL class through teacher self-observation-introspection. *International Journal of Bilingual Education and Bilingualism*, 10(5), 625–646.
- Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge, U.K: Press Syndicate of the University of Cambridge.
- Coxhead, A. (1998). *An academic word list*. School of Linguistics and Applied Language Studies, Victoria University of Wellington.
- Coyle, D. (1999). Theory and planning for effective classrooms: Supporting students in content and language integrated learning contexts. In J. Masih (Ed.) *Learning Through a Foreign Language* (pp. 46-62). London: Centre for Information on Language Teaching and Research.

- Coyle, D. (2007). Content and language integrated learning: Towards a connected research agenda for CLIL pedagogies. *International journal of bilingual education and bilingualism*, 10(5), 543-562.
- Coyle, D. (2008). CLIL – a pedagogical approach. In N. van Deusen-Scholl, & N. Hornberger (Eds.), *Encyclopedia of Language and Education* (2nd ed. pp. 97-111). New York: Springer.
- Coyle, D. (2010). Forward. In Y. Ruiz de Zarobe & D. Lasagabaster (Eds.), *CLIL in Spain: Implementation, results and teacher training* (pp. vii-viii). Newcastle upon Tyne: Cambridge Scholars Publishing.
- Coyle, D., Hood, P., & Marsh, D. (2010). *CLIL: Content and Language Integrated Learning*. Cambridge: Cambridge University Press.
- Crookes, G., & Schmidt, R. W. (1991). Motivation: Reopening the research agenda. *Language Learning*, 41(4), 469-512.
- Cross, R. (2012). Creative in finding creativity in the curriculum: The CLIL second language classroom. *The Australian Educational Researcher*, 39(4), 431-445.
- Csizér, K. (2017). Motivation in the L2 classroom. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 418-432). New York: Routledge.
- Csizér, K. (2019). The L2 motivational self system. In M. Lamb, K. Csizér, A. Henry & S. Ryan (Eds.), *The Palgrave Handbook of Motivation for Language Learning* (pp. 71-93). Cham: Palgrave Macmillan.
- Csizér, K., & Dörnyei, Z. (2005a). Language learners' motivational profiles and their motivated learning behaviour. *Language Learning*, 55(4), 613–59.
- Csizér, K., & Dörnyei, Z. (2005b). The internal structure of language learning motivation and its relationship with language choice and learning effort. *The modern language journal*, 89(1), 19-36.
- Csizér, K., & Kormos, J. (2009). Learning experiences, selves and motivated learning behaviour: A comparative analysis of structural models for Hungarian secondary and university learners of English. In Z. Dörnyei, & E. Ushioda (Eds.), *Motivation, Language Identity and the L2 Self* (pp. 98-119). Bristol: Multilingual Matters.

- Csizér, K., & Lukács, G. (2010). The comparative analysis of motivation, attitudes and selves: The case of English and German in Hungary. *System*, 38(1), 1-13.
- Csizér, K., & Tankó, G. (2017). English majors' self-regulatory control strategy use in academic writing and its relation to L2 motivation. *Applied Linguistics*, 38(3), 386-404.
- Culligan, B. (2015). A comparison of three test formats to assess word difficulty. *Language Testing*, 32(4), 503–520.
- Cummins, J. (1979) Cognitive/academic language proficiency, linguistic interdependence, the optimum age question and some other matters. *Working Papers on Bilingualism*, 19, 121-129.
- Cummins, J. (2013). BICS and CALP. In P. Robinson (Ed.), *The Routledge Encyclopedia of Second Language Acquisition* (pp. 65-66). New York: Routledge.
- Dale, E. (1965). Vocabulary measurement: Techniques and major findings. *Elementary English*, 42, 895–901, 948.
- Daller, H., & Phelan, D. (2006). The C-test and TOEIC® as measures of students' progress in intensive short courses in EFL. In R. Grotjahn (Ed.), *Der C-Test: Theorie, Empirie, Anwendungen/The C-Test: Theory, empirical research, applications* (pp. 101-119). Frankfurt am Main: Peter Lang.
- Daller, H., & Xue, H. (2007). Lexical richness and the oral proficiency of Chinese EFL students. In H. Daller, J. Milton and J. Treffers-Daller (Eds.), *Modelling and Assessing Vocabulary Knowledge* (pp. 150-164). Cambridge, UK: Cambridge University Press.
- Daller, H., Milton, J., & Treffers-Daller J. (2007). Editors' introduction: Conventions, terminology and an overview of the book. In H. Daller, J. Milton, & J. Treffers-Daller (Eds.), *Modelling and assessing vocabulary knowledge* (pp. 1-32). Cambridge, UK: Cambridge University Press.
- Daller, H., van Hout, R., & Daller-Treffers, J. (2003). Lexical richness in the spontaneous speech of bilinguals. *Applied Linguistics*, 24, 197-222.
- Daller, M. H., & Xue, H. (2009). Vocabulary Knowledge and Academic Success: A Study of Chinese Students in UK Higher Education. In B. Richards, M. H. Daller, D. D. Malvern, P. Meara, J. Milton & J. Treffers-Daller (Eds.), *Vocabulary studies in*

- first and second language acquisition* (pp. 179-193). London: Palgrave Macmillan.
- Dalton-Puffer, C. (2007). *Discourse in content and language integrated learning (CLIL) classrooms*. Amsterdam: John Benjamins.
- Dalton-Puffer, C. (2008). Outcomes and processes in content and language integrated learning (CLIL): Current research from Europe. In W. Delanoy & L. Volkman (Eds.), *Future perspectives for English language teaching* (pp. 139-157). Heidelberg, Germany: Carl Winter.
- Dalton-Puffer, C. (2011). Content-and-language integrated learning: From practice to principles? *Annual Review of applied linguistics*, 31, 182-204.
- Dalton-Puffer, C., & Smit, U. (2007). Introduction. In C. Dalton-Puffer & U. Smit (Eds.), *Empirical Perspectives on CLIL Classroom Discourse* (pp. 7-23). Frankfurt: Peter Lang.
- Dalton-Puffer, C., & Smit, U. (2013). Content and language integrated learning: A research agenda. *Language Teaching*, 46(4), 545–559.
- Dalton-Puffer, C., Llinares, A., Lorenzo, F., & Nikula, T. (2014). “You can stand under my umbrella”: Immersion, CLIL and bilingual education. A response to Cenoz, Genesee & Gorter (2013). *Applied Linguistics*, 35(2), 213-218.
- Dalton-Puffer, C., Nikula, T., & Smit, U. (2010). *Language use and language learning in CLIL classrooms: Current findings and contentious issues*. Amsterdam: John Benjamins.
- de Bot, K., Lowie, W. and Verspoor, M. (2007). A Dynamic Systems Theory approach to second language acquisition. *Bilingualism, Language and Cognition*, 10(1), 7–21.
- DeCarrico, J. S. (2001). Vocabulary learning and teaching. In M. Celce-Murcia (Ed.). *Teaching English as a second or foreign language* (pp. 285-299). Boston: Heinle and Heinle.
- de la Maya Retamar, M. (2016). *El desarrollo de la competencia léxica en la adquisición del francés como segunda lengua extranjera en la ESO* (Unpublished doctoral dissertation). Universidad de Extremadura, Spain.

- De Smet, A., Mettwie, L., Galand, B., Hiligsmann, P., & van Mensel, L. (2018). Classroom anxiety and enjoyment in CLIL and non-CLIL: Does the target language matter? *Studies in Second Language Learning and Teaching*, 8, 47-72.
- De Smet, A., Mettwie, L., Hiligsmann, P., Galand, B., & van Mensel, L. (2019). Does CLIL shape language attitudes and motivation? Interactions with target languages and instruction levels. *International Journal of Bilingual Education and Bilingualism*. <https://doi.org/10.1080/13670050.2019.1671308>
- Dearden, J. (2014). *English as a medium of instruction-a growing global phenomenon*. London: British Council.
- Deci, E. L., & Ryan, R. M. (1985). The general causality orientations scale: Self-determination in personality. *Journal of research in personality*, 19(2), 109-134.
- Deller, S., & Price, C. (2007). *Teaching other subjects through English*. Oxford: Oxford University Press.
- Dickinson, L. (1995). Autonomy and motivation a literature review. *System*, 23(2), 165-174.
- Diéz Prados, M. (2010). Gender and L1 influence on EFL learners' lexicon. In R. M. Jiménez Catalán (Ed.), *Gender Perspectives on Vocabulary in Foreign and Second languages* (pp. 44-73). New York: Palgrave Macmillan.
- Dimitrijević, N. (1969). *Lexical availability*. Heidelberg, Germany: Julius Gross Verlag.
- Dóczi, B., & Kormos, J. (2016). *Longitudinal developments in vocabulary knowledge and lexical organization*. Oxford: Oxford University Press.
- Doiz, A., Lasagabaster, D., & Sierra, J. M. (2014a). CLIL and motivation- The effect of individual and contextual variables. *The Language Learning Journal*, 42(2), 209-224.
- Doiz, A., Lasagabaster, D., & Sierra, J. M. (2014b). Giving voice to the students: What (de) motivates them in CLIL classes? In D. Lasagabaster, A. Doiz, & J. M. Sierra (Eds.), *Motivation and foreign language learning: From theory to practice* (pp. 117-138). Amsterdam: John Benjamins Publishing Company.
- Dörnyei, Z. (1994a). Understanding L2 motivation: On with the challenge! *The Modern Language Journal*, 78(4), 515-523.

- Dörnyei, Z. (1994b). Motivation and motivating in the foreign language classroom. *The Modern Language Journal*, 78(3), 273-284.
- Dörnyei, Z. (2001). New themes and approaches in second language motivation research. *Annual Review of Applied Linguistics*, 21(1), 43-61.
- Dörnyei, Z. (2002) The motivational basis of language learning tasks. In P. Robinson (Ed.), *Individual Differences and Instructed Language Learning* (pp. 137–158). Amsterdam: John Benjamins.
- Dörnyei, Z. (2003). *Questionnaires in second language research: Constructing, administering, and processing*. London: Lawrence Erlbaum Associates.
- Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Mahwah, NJ: Lawrence Erlbaum.
- Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford: Oxford University Press.
- Dörnyei, Z. (2009a). *The Psychology of Second Language Acquisition*. Oxford: Oxford University Press.
- Dörnyei, Z. (2009b). The L2 motivational self system. Motivation, Language Identity and the L2 Self, In Z. Dörnyei, & E. Ushioda (Eds.), *Motivation, Language Identity and the L2 Self* (pp. 9-42). Bristol: Multilingual Matters.
- Dörnyei, Z. (2010). The relationship between language aptitude and language learning motivation: Individual differences from a dynamic systems perspective. In E. Macaro (Ed.), *Continuum Companion to Second Language Acquisition* (pp. 247-267). London: Continuum.
- Dörnyei, Z. (2014). *The Psychology of the Language Learner: Individual Differences in Second Language Acquisition*. London: Routledge.
- Dörnyei, Z. (2019). From integrative motivation to directed motivational currents: The evolution of the understanding of L2 motivation over three decades. In M. Lamb, K. Csizér, A. Henry & S. Ryan (Eds.), *The Palgrave Handbook of Motivation for Language Learning* (pp. 39-69). Cham: Palgrave Macmillan.

- Dörnyei, Z., & Al-Hoorie, A. H. (2017). The motivational foundation of learning languages other than Global English: Theoretical issues and research directions. *The Modern Language Journal*, 101(3), 455-468.
- Dörnyei, Z., & Chan, L. (2013). Motivation and vision: An analysis of future L2 self images, sensory styles, and imagery capacity across two target languages. *Language Learning*, 63(3), 437-462.
- Dörnyei, Z., & Csizér, K. (2002). Some dynamics of language attitudes and motivation: Results of a longitudinal nationwide survey. *Applied Linguistics*, 23(4), 421-462.
- Dörnyei, Z., & Csizér, K. (2005). The effects of intercultural contact and tourism on language attitudes and language learning motivation. *Journal of Language and Social Psychology*, 24(4), 327-357.
- Dörnyei, Z., & Csizér, K. (2012). How to Design and Analyze Surveys in Second Language Acquisition. In A. Mackey & S. M. Gass (Eds.), *Research Methods in Second Language Acquisition: A Practical Guide* (pp. 74–94). Oxford: Wiley-Blackwell.
- Dörnyei, Z., & Kormos, J. (2000). The role of individual and social variables in oral task performance. *Language Teaching Research*, 4(3), 275-300.
- Dörnyei, Z., & Ottó, I. (1998). Motivation in action: A process model of L2 motivation. *Working Papers in Applied Linguistics* (Thames Valley University, London), 4, 43-69.
- Dörnyei, Z., & Ryan, S. (2015). *The psychology of the language learner revisited*. London: Routledge.
- Dörnyei, Z., & Taguchi, T. (2009). *Questionnaires in second language research: Construction, administration, and processing*. New York: Routledge.
- Dörnyei, Z., & Ushioda, E. (2009). Motivation, language identities and the L2 self: a theoretical overview. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, Language and the L2 Self* (pp. 1–8). Bristol: Multilingual Matters.
- Dörnyei, Z., & Ushioda, E. (2013). *Teaching and researching: Motivation* (2nd ed.). London: Routledge.

- Dörnyei, Z., & Ushioda, E. (Eds.) (2009). *Motivation, Language Identity and the L2 Self*. Bristol: Multilingual Matters.
- Dörnyei, Z., Csizér, K., & Németh, N. (2006). *Motivation, language attitudes and globalisation: A Hungarian perspective*. Clevedon: Multilingual Matters.
- Dörnyei, Z., MacIntyre, P. D., & Henry, A. (2014). Introduction: Applying complex dynamic systems principles to empirical research on L2 motivation. In Z. Dörnyei, P. D. MacIntyre, & A. Henry (Eds.), *Motivational Dynamics in Language Learning* (pp. 1-7). Bristol: Multilingual Matters.
- Dörnyei, Z., MacIntyre, P. D., & Henry, A. (Eds.) (2014). *Motivational Dynamics in Language Learning*. Bristol: Multilingual Matters.
- East, M. (2006). The impact of bilingual dictionaries on lexical sophistication and lexical accuracy in tests of L2 writing proficiency: A quantitative analysis. *Assessing Writing, 11*(3), 179-197.
- Eckes, T., & Grotjahn, R. (2006). A closer look at the construct validity of C-tests. *Language Testing, 23*, 290-325.
- Ehrlich, S. (1997). Gender as social practice: Implications for second language acquisition. *Studies in Second Language Acquisition, 19*(4), 421-446.
- Ehrman, M., & Oxford, R. (1989). Effects of sex differences, career choice, and psychological type on adult language learning strategies. *The Modern Language Journal, 73*(1), 1-13.
- Ellis, N. C. (2007). Dynamic systems and SLA: The wood and the trees. *Bilingualism, Language and Cognition, 10*(1), 23-25.
- Ellis, R. (1994). *The study of second language acquisition*. Oxford University.
- Ellis, R. (2001). Introduction: Investigating form-focused instruction. *Language Learning, 51*, 1-46.
- Ellis, R., & Shintani, N. (2013). *Exploring language pedagogy through second language acquisition research*. New York: Routledge.
- Engber, C. A. (1995). The relationship of lexical proficiency to the quality of ESL compositions. *Journal of second language writing, 4*(2), 139-155.

- Erigna, D. (1974). Enseigner, c'est choisir: vocabulaire-verwerving. *Levende Talen*, 306, 260-267.
- Eriksson, M., Marschik, P. B., Tulviste, T., Almgren, M., Pérez Pereira, M., Wehberg, S., Marjanovič-Umek, L., Gayraud, F., Kovacevic, M., & Gallego, C. (2012). Differences between girls and boys in emerging language skills: Evidence from 10 language communities. *British Journal of Developmental Psychology*, 30(2), 326-343.
- Etchell, A., Adhikari, A., Weinberg, L. S., Choo, A. L., Garnett, E. O., Chow, H. M., & Chang, S. E. (2018). A systematic literature review of sex differences in childhood language and brain development. *Neuropsychologia*, 114, 19-31.
- European Commission (2012). *Europeans and their languages. Special Eurobarometer 386*. Brussels: European Commission.
- Eurydice (2006). *Content and Language Integrated Learning (CLIL) at School in Europe*. Brussels: European Commission.
- Eurydice (2017). *Key data on teaching languages at school in Europe – 2017 Edition*. Eurydice Report. Luxembourg: Publications Office of the European Union.
- Eyckmans, J., van de Velde, H., van Hout, R. and Boers, F. (2007). Learners' response behaviour in Yes/No vocabulary tests. In H. Daller, J. Milton and J. Treffers-Daller (Eds.), *Modelling and Assessing Vocabulary Knowledge* (pp. 59-76). Cambridge, UK: Cambridge University Press.
- Fan, M. (2000). How big is the gap and how to narrow it? An investigation into the active and passive vocabulary knowledge of L2 learners. *RELC Journal*, 31(2), 105-119.
- Farhady, H., & Jamali, F. (2006). Varieties of C-test as measures of general language proficiency. In H. Farhadi (Ed.), *Twenty-five years of living with applied linguistics: A collection of articles* (pp. 287-302). Oxford: B. Blackwell.
- Fernández Fontecha, A. (2009). Spanish CLIL: Research and official actions. In Y. Ruiz de Zarobe & R. M. Jiménez Catalán (Eds.), *Content and language integrated learning Evidence from research in Europe* (pp. 3-21). Bristol: Multilingual Matters.

- Fernández Fontecha, A. (2010). Gender and motivation in EFL vocabulary production. In R. M. Jiménez Catalán (Ed.), *Gender Perspectives on Vocabulary in Foreign and Second languages* (pp. 93–117). New York: Palgrave Macmillan.
- Fernández Fontecha, A. (2014a). Motivation and gender effect in receptive vocabulary learning: An exploratory analysis in CLIL primary education. *Latin American Journal of Content & Language Integrated Learning*, 7(2), 27-49.
- Fernández Fontecha, A. (2014b). Individual and contextual variables in EFL receptive vocabulary: analysis of gender effect and motivation of CLIL primary graders. *ODISEA. Revista de Estudios Ingleses*, 15, 57-72.
- Fernández Fontecha, A. (2014c). Receptive vocabulary knowledge and motivation in CLIL and EFL. *Revista de Lingüística y Lenguas Aplicadas*, 9, 23-32.
- Fernández Fontecha, A. (2015). Motivation and vocabulary breadth in CLIL and EFL contexts. Different age, same time of exposure. *Complutense Journal of English Studies*, 23, 79-96.
- Fernández Fontecha, A., & Canga Alonso, A. (2014). A preliminary study on motivation and gender in CLIL and non-CLIL types of instruction. *International Journal of English Studies*, 14(1), 21-36.
- Fernández Orío, S., & Jiménez Catalán, R. M. (2015). Lexical availability of EFL learners at the end of Spanish Secondary Education: The effect of language program and prompt. *ES Revista de Filología Inglesa*, 36, 103-127.
- Fernández Smith, G., Sánchez-Saus Laserna, M., & Escoriza Morerer, L. (2012). Studies on Lexical Availability: The Current Situation and Some Future Prospects. In B. Eizaga Rebollar (Ed.), *Studies in Linguistics and Cognition* (pp. 35-57). Berlin: Peter Lang.
- Fernández-Sanjurjo, J., Fernández-Costales, A., & Arias Blanco, J. M. (2019). Analysing students' content-learning in science in CLIL vs. non-CLIL programmes: Empirical evidence from Spain. *International Journal of Bilingual Education and Bilingualism*, 22(6), 661-674.
- Ferreira, R. A., Garrido Moscoso, J. I., & Guerra Rivera, A. (2019). Predictors of lexical availability in English as a second language. *Onomázein: Revista de lingüística, filología y traducción de la Pontificia Universidad Católica de Chile*, 46, 18-34.

- Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). London: Sage publications.
- Fitzpatrick, T. (2007). Productive vocabulary tests and the search for concurrent validity. In H. Daller, J. Milton and J. Treffers-Daller (Eds.), *Modelling and Assessing Vocabulary Knowledge* (pp. 116-132). Cambridge, UK: Cambridge University Press.
- Fitzpatrick, T., & Clenton, J. (2010). The challenge of validation: Assessing the performance of a test of productive vocabulary. *Language Testing*, 27(4), 537-554.
- Fitzpatrick, T., & Meara, P. (2004). Exploring the validity of a test of productive vocabulary. *Vigo International Journal of Applied Linguistics*, 1, 55-74.
- Fontecha, A. F. (2010). First steps of CLIL in a Spanish monolingual community: The case of La Rioja. In Y. Ruiz de Zarobe & D. Lasagabaster (Eds.), *CLIL in Spain: Implementation, results and teacher training* (pp. 79-94). Newcastle upon Tyne: Cambridge Scholars Publishing.
- Francis, W. N., & Kučera, H. (1982). *Frequency analysis of English usage*. Boston, MA: Houghton Mifflin.
- Friedman, D. A. (2012). How to Collect and Analyze Qualitative Data. In A. Mackey & S. M. Gass (Eds.), *Research methods in second language acquisition: A practical guide* (pp.180-200). John Wiley & Sons.
- Frigols Martín, M. J. (2008). CLIL implementation in Spain: An approach to different models. In C. M. Coonan (Ed.), *CLIL e l'apprendimento delle lingue. Le sfide del nuovo ambiente di apprendimento, Venezia, settembre 2008. Venezia: Libreria Editrice Cafoscarina, 2008* (pp. 221-232). Venezia, Libreria Editrice Cafoscarina.
- Fryer, L. K., Larson-Hall, J., & Stewart, J. (2018). Quantitative methodology. In A. Phakiti, P. de Costa, L. Plonsky, & S. Starfield (Eds.), *The Palgrave handbook of applied linguistics research methodology* (pp. 55-77). London: Palgrave Macmillan.
- Fukada, Y., Falout, J., Fukuda, T., & Murphey, T. (2019). Motivational Group Dynamics in SLA: The Interpersonal Interaction Imperative. In M. Lamb, K. Csizér, A. Henry & S. Ryan (Eds.), *The Palgrave Handbook of Motivation for Language Learning* (pp. 307-325). Cham: Palgrave Macmillan, Cham.

- Fukada, Y., Murphey, T., Falout, J., & Fukuda, T. (2017). Essential motivational group dynamics: A 3-year panel study. In R. Breeze & C. S. Guinda (Eds.), *Essential competencies for English-medium university teaching* (pp. 249–266). Switzerland: Springer.
- Gabriella Morreale, S. (2011). *The relationship between study abroad and motivation, attitude and anxiety in university students learning a foreign language*. (Unpublished doctoral dissertation). Wayne State University, Michigan.
- Gallardo del Puerto, F., & Martínez Adrián, M. (2014). The incidence of previous foreign language contact in a lexical availability task: A study of senior learners. In R. M. Jiménez Catalán (Ed.), *Lexical Availability in English and Spanish as a Second Language* (pp. 53-68). Dordrecht: Springer.
- Gallardo del Puerto, F., Gómez Lacabex, E., & García Lecumberri, M. L. (2009). Testing the effectiveness of content and language integrated learning in foreign language contexts: The assessment of English pronunciation. In Y. Ruiz de Zarobe & R. M. Jiménez Catalán (Eds.), *Content and language integrated learning Evidence from research in Europe* (pp. 63-80). Bristol: Multilingual Matters.
- Gallardo-del-Puerto, F., & Blanco-Suárez, Z. (2021). Foreign Language Motivation in Primary Education students: The effects of additional CLIL and gender. *Journal of Immersion and Content-based Language Education*, 9(1), 58-84.
- Gardner, D. (2007). Validating the construct of word in applied corpus-based vocabulary research: A critical survey. *Applied Linguistics*, 28(2), 241-265.
- Gardner, R. C. (1979). Social psychological aspects of second language acquisition. In H. Giles & R. St. Clair (Eds.), *Language and Social Psychology* (pp. 193-200). Oxford: Blackwell.
- Gardner, R. C. (1985) *Social Psychology and Second Language Learning: The Role of Attitudes and Motivation*. London: Edward Arnold.
- Gardner, R. C. (2001). Integrative Motivation and Second Language Acquisition. In Z. Dörnyei & R. Schmidt (Eds.), *Motivation and second language acquisition* (pp. 1-20). Honolulu, HI: University of Hawaii Press.
- Gardner, R. C. (2006). The socio-educational model of second language acquisition: A research paradigm. *Eurosla Yearbook*, 6(1), 237-260.

- Gardner, R. C. (2010). *Motivation and second language acquisition: The socio-educational model*. New York: Peter Lang.
- Gardner, R. C. (2019). From The Socio-educational Model of Second Language Acquisition. In M. Lamb, K. Csizér, A. Henry & S. Ryan (Eds.), *The Palgrave Handbook of Motivation for Language Learning* (pp. 21-37). Cham: Palgrave Macmillan.
- Gardner, R. C., & Lambert, W. E. (1972). *Attitudes and motivation in second-language learning*. Rowley, MA: Newbury House Publishers
- Gardner, R. C., & MacIntyre, P. D. (1993). On the measurement of affective variables in second language learning. *Language Learning*, 43(2), 157-194.
- Gardner, R. C., & Smythe, P. C. (1974). The integrative motive in second-language acquisition. *Bilingualism, Biculturalism and Education*, 1, 31–45.
- Gardner, R. C., & Tremblay, P. F. (1994a). On motivation, research agendas and theoretical frameworks. *The Modern Language Journal*, 78(3), 359–368.
- Gardner, R. C., & Tremblay, P. F. (1994b). On motivation: Measurement and conceptual considerations. *The Modern Language Journal*, 78(4), 524–527.
- Gass, S. M. (2013). *Second Language Acquisition: An Introductory Course* (4th ed.). London: Routledge.
- Gass, S., & Varonis, E. (1986). Sex differences in NNS/NNS interactions. In R. Day (Ed.), *Talking to Learn: Conversation in Second Language Acquisition*. Rowley, MA: Newbury House.
- Geoghegan, L. (2018). International posture, motivation and identity in SA. En C. Pérez-Vidal, S. López-Serrano, J. Ament and D. Thomas Wilhelm (Eds.), *Learning Context Effects: Study Abroad, Formal Instruction and International Immersion Classrooms* (pp. 215-253). Berlin: Language Science Press.
- Geoghegan, L., & Pérez-Vidal, C. (2019). English as a Lingua Franca, motivation and identity in Study Abroad. En M. Howard (Ed.), *Study Abroad, Second Language Acquisition and Interculturality: Contemporary Perspectives* (pp. 103-135). Bristol: Multilingual Matters.

- Ghaith, G. (2003). The relationship between forms of instruction, achievement and perceptions of classroom climate. *Educational Research*, 45, 89–93.
- Ghazvini, S. D., & Khajepour, M. (2011). Attitudes and motivation in learning English as second language in high school students. *Procedia-Social and Behavioral Sciences*, 15, 1209-1213.
- Gierlinger, E. M., & Wagner, T. A. (2016). The more the merrier—revisiting CLIL-based vocabulary growth in secondary education. *Latin American Journal of content & language integrated learning*, 9(1), 37-63.
- Giles, H. and Byrne, J. L. (1982) An intergroup approach to second language acquisition. *Journal of Multilingual and Multicultural Development*, 3, 17–40.
- Gobierno de La Rioja (2013). *Convocatoria de proyecto de competencia lingüística integrada en centros (CLIC). Curso 2013/2014*. Retrieved from <https://www.larioja.org/larioja-client/cm/edu-innovacion-form/images?idMmedia=609148>
- Gobierno de la Rioja (n.d.). Retrieved from <https://www.larioja.org/edu-innovacion-form/es/lenguas-extranjeras/bilinguismo>
- Gobierno de La Rioja. (2019). *Relación definitiva de centros y docentes participantes en el Proyecto de Innovación Lingüística en Centros (PILC) durante el curso 2019-2020*. Retrieved from <https://www.larioja.org/edu-innovacion-form/es/lenguas-extranjeras/bilinguismo/proyectos-innovacion-linguistica-centros.ficheros/1163328-PILC%20Definitivo%202019-20%20f.pdf>
- Goh, H. H. (2017). *Mandarin Competence of Chinese-English Bilingual Preschoolers*. Singapore: Springer.
- Goris, J. A., Denessen, E. J. P. G., & Verhoeven, L. T. W. (2019). Effects of content and language integrated learning in Europe A systematic review of longitudinal experimental studies. *European Educational Research Journal*, 18(6), 675-698.
- Gougenheim, G., R. Michéa, P. Rivenc and A. Sauvageot (1964). *L'élaboration du Français Fondamental (1er degré). Étude sur l'élaboration d'un vocabulaire et d'une grammaire de base*. Paris: Didier
- Graddol, D. (2006). *English next*. London: British Council.

- Graham, K. M., Choi, Y., Davoodi, A., Razmeh, S., & Dixon, L. Q. (2018). Language and Content Outcomes of CLIL and EMI: A Systematic Review. *Latin American Journal of Content and Language Integrated Learning*, 11(1), 19-37.
- Graham, S. (1994) Classroom motivation from an attributional perspective. In H.F. O'Neil & M. Drillings (Eds), *Motivation: Theory and Research* (pp. 31-48). Hillsdale, NJ: Lawrence Erlbaum.
- Grotjahn, R., & Stemmer, B. (2002). C-Tests and language processing. In J. A. Coleman, R. Grotjahn & U. Raatz (Eds.), *University Language Testing and The C-Test* (pp. 115-130). Bochum: AKS-Verlag.
- Grujić, T., & Danilović, J. (2012, April). *The Value of C-Tests in English Language Testing and Teaching*. Paper presented at Multidisciplinary Conferences Language, literature, values, Faculty of Philosophy, University of Niš, Serbia.
- Guadamillas Gómez, M. V., & Alcaraz Mármol, G. (2018). Legislación en enseñanza bilingüe: análisis en el marco de educación primaria en España [Bilingual Legislation. Analysing the Legal Framework for Primary Education in Spain]. *MULTIárea. Revista de Didáctica*, 9, 82-103.
- Guerrero, M. (2015). Motivation in second language learning: A historical overview and its relevance in a public high school in Pasto, Colombia. *HOW*, 22(1), 95-106.
- Guilloteaux, M. J., & Dörnyei, Z. (2008). Motivating language learners: A classroom-oriented investigation of the effects of motivational strategies on student motivation. *TESOL Quarterly*, 42(1), 55-77.
- Guiraud, P. (1954). *Les caractères statistiques du vocabulaire*. Paris: Presses universitaires de France.
- Hadfield, J., & Dörnyei, Z. (2013). *Motivating learning*. Harlow: Longman.
- Halliday, T. C. (1985). Knowledge mandates: collective influence by scientific, normative and syncretic professions. *British Journal of Sociology*, 36, 421-447.
- Halpern, D. F (2000). *Sex differences in cognitive abilities* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Harrington, M. (2018). *Lexical Facility*. London: Palgrave Macmillan.

- Hartmann, G. W. (1946). Further evidence on the unexpected large size of recognition vocabularies among college students. *Journal of Educational Psychology*, 37(7), 436.
- Hartshorne, J. K., & Ullman, M. T. (2006). Why girls say 'holded' more than boys. *Developmental Science*, 9(1), 21-32.
- Hastings, A. J. (2002). In defense of C-testing. In R. Grotjahn (Ed.), *Der C-Test: Theoretische Grundlagen und praktische Anwendungen* (pp. 11-25). Bochum: AKS-Verlag.
- Healy, A. F., & Sherrod, N. B. (1994, November). *The/Thee pronunciation distinction: A local model of linguistic categories*. Paper presented at the 35th Annual Meeting of the Psychonomic Society, St. Louis, MO.
- Heatley, A., Nation, I. S. P., & Coxhead, A. (2002). RANGE and FREQUENCY programs [Computer software]. *Wellington, New Zealand: Victoria University of Wellington*.
- Heckhausen, H., & Kuhl, J. (1985) From wishes to action: The dead ends and short cuts on the long way to action. In M. Frese & J. Sabini (Eds), *Goal-Directed Behaviour: The Concept of Action in Psychology* (pp. 134-160). Hillsdale, NJ: Lawrence Erlbaum.
- Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second Language Acquisition*, 21(2), 303-317.
- Henry, A. (2009). Gender differences in compulsory school pupils' L2 self-concepts: A longitudinal study. *System*, 37(2), 177-193.
- Henry, A. (2010). Contexts of possibility in simultaneous language learning: Using the L2 Motivational Self System to assess the impact of global English. *Journal of Multilingual and Multicultural development*, 31(2), 149-162.
- Henry, A. (2011). Examining the impact of L2 English on L3 selves: A case study. *International Journal of Multilingualism*, 8(3), 235-255.
- Henry, A. (2014a). The Motivational Effects of Crosslinguistic Awareness: Developing Third Language Pedagogies to Address the Negative Impact of the L2 on the L3 Self-Concept. *Innovation in Language Learning and Teaching*, 8(1), 1-19.

- Henry, A. (2014b). The Dynamics of L3 Motivation: A Longitudinal Interview/Observation Based Study. In Z. Dörnyei, P. D. MacIntyre, & A. Henry (Eds.), *Motivational Dynamics in Language Learning* (pp. 315-342). Bristol: Multilingual Matters.
- Henry, A. (2017). L2 motivation and multilingual identities. *The Modern Language Journal*, 101(3), 548-565.
- Heras, A., & Lasagabaster, D. (2015). The impact of CLIL on affective factors and vocabulary learning. *Language Teaching Research*, 19(1), 70-88.
- Hernández Muñoz, N. (2010). Social Aspects of Oral and Written Lexical Production in Spanish. *SKY Journal of Linguistics*, 23, 101-123.
- Hernández, T. A. (2010). The relationship among motivation, interaction, and the development of second language oral proficiency in a study-abroad context. *The Modern Language Journal*, 94(4), 600-617.
- Hernández-Muñoz, N., Izura, C., & Ellis, A. W. (2006). Cognitive aspects of lexical availability. *European Journal of Cognitive Psychology*, 18(5), 730-755.
- Hessel, G. (2015). From vision to action: Inquiring into the conditions for the motivational capacity of ideal second language selves. *System*, 52, 103-114.
- Higgins, E. T. (1987). Self-discrepancy: a theory relating self and affect. *Psychological Review*, 94(3), 319.
- Hilton, H. (2008). The link between vocabulary knowledge and spoken L2 fluency. *Language Learning Journal*, 36(2), 153-166.
- Warshani Himanshi Pilimatalawwe, W. (2015). *A Study of Second Language Identity and Motivation among Undergraduates in Sri Lankan Universities*. (Unpublished Master's thesis). St. Cloud State University, Minnesota.
- Hindmarsh, R. (1980). *Cambridge English lexicon*. Cambridge: Cambridge University Press.
- Hummel, K. M. (2013). *Introducing second language acquisition: Perspectives and practices*. Blackwell: John Wiley & Sons.
- Iglesias Diéguez, K., & Martínez-Adrián, M. (2017). The influence of CLIL on receptive vocabulary: A preliminary study. *Journal of English Studies*, 15, 107-134.

- Instituto Nacional de Estadística (2021). *Población por comunidades y ciudades autónomas y sexo 2021*. Retrieved from <https://www.ine.es/jaxiT3/Datos.htm?t=2853>
- Ioannou Georgiou, S. (2012). Reviewing the puzzle of CLIL. *ELT journal*, 66(4), 495-504.
- Iqbal, S., & Yongbing, L. (2018). L2 Anxiety Influencing Self-Reported Language Proficiency: A Quantitative Study on Pakistani ESL Arts Students. *European Journal of English Language Teaching*, 4(2), 78-89.
- Irie, K., & Ryan, S. (2015). Study abroad and the dynamics of change in learner L2 self-concept. In Z. Dörnyei, P. D. MacIntyre, & A. Henry (Eds.), *Motivational Dynamics in Language Learning Motivational Dynamics in Language Learning* (pp. 343–366). Bristol, UK: Multilingual Matters.
- Isabelli-García, C. (2006). Study abroad social networks, motivation and attitudes: Implications for second language acquisition. *Language Learners in Study Abroad Contexts*, 15, 231-258.
- Islam, M., Lamb, M., & Chambers, G. (2013). The L2 motivational self system and national interest: A Pakistani perspective. *System*, 41(2), 231-244.
- James, M. (1996). *Improving second language reading comprehension: A computer-assisted vocabulary development approach*. (Unpublished doctoral dissertation). University of Hawaii, Monoa.
- Järvinen, H. (2005). Language learning in content-based instruction. In A. Housen & M. Pierrard (Eds), *Investigations in second language acquisition* (pp. 433-546). Berlin: Mouton de Gruyter.
- Jenkins, J. (2002). A sociolinguistically based, empirically researched pronunciation syllabus for English as an international language. *Applied Linguistics*, 23(1), 83-103.
- Jiménez Catalán, R. M. (2002). El concepto de competencia léxica en los estudios de aprendizaje y enseñanza de segundas lenguas. *Atlantis*, 24(2), 149-162.
- Jiménez Catalán, R. M. (2003). Sex differences in L2 vocabulary learning strategies. *International Journal of Applied Linguistics*, 13(1), 54-77.

- Jiménez Catalán, R. M. (2005). Linking gender and second language education in a database. *Cauce*, 28, 205-218.
- Jiménez Catalán, R. M. (2010). Gender tendencies in EFL across vocabulary tests. In R. M. Jiménez Catalán (Ed.), *Gender Perspectives on Vocabulary in Foreign and Second languages* (pp. 117–139). New York: Palgrave Macmillan.
- Jiménez Catalán, R. M. (2014). Preface. In R. M. Jiménez Catalán (Ed.), *Lexical Availability in English and Spanish as a Second Language* (pp. v-ix). Dordrecht: Springer.
- Jiménez Catalán, R. M. (Ed.) (2010). *Gender Perspectives on Vocabulary in Foreign and Second languages*. New York: Palgrave Macmillan.
- Jiménez Catalán, R. M., & Agustín Llach, M. P. (2017). CLIL or Time? Lexical profiles of CLIL and non-CLIL EFL learners. *System*, 66, 87-99.
- Jiménez Catalán, R. M., & Canga Alonso, A. (2019). El léxico disponible en inglés de chicos y chicas españoles adolescentes. *ELIA: Estudios de Lingüística Inglesa Aplicada*, (19), 157-176.
- Jiménez Catalán, R. M., & Dewaele, J. M. (2017). Lexical availability of young Spanish EFL learners: Emotion words versus non-emotion words. *Language, Culture and Curriculum*, 30(3), 283-299.
- Jiménez Catalán, R. M., & Fernández Fontecha, A. (2015). Lexical phrases in compositions by CLIL and non-CLIL EFL learners. In D. Marsh, M. L. Pérez Cañado & J. Ráez Padilla (Eds.), *CLIL in Action: Voices from the Classroom* (pp. 82-97). Newcastle upon Tyne: Cambridge Scholars Publishing.
- Jiménez Catalán, R. M., & Fernández Fontecha, A. (2019). Lexical Availability Output in L2 and L3 EFL Learners: Is There a Difference? *English Language Teaching*, 12(2), 77-87.
- Jiménez Catalán, R. M., & Fitzpatrick, T. (2014). Frequency profiles of EFL learners' lexical availability. In R. M. Jiménez Catalán (Ed.), *Lexical Availability in English and Spanish as a Second Language* (pp. 83-100). Dordrecht: Springer.
- Jiménez Catalán, R. M., & Moreno Espinosa, S. (2005). Using Lex30 to measure the L2 productive vocabulary of Spanish primary learners of EFL. *Vigo International Journal of Applied Linguistics*, 2, 27-44.

- Jiménez Catalán, R. M., & Ojeda Alba, J. (2009a). Girls' and boys' lexical availability in EFL. *ITL International Journal of Applied Linguistics*, 158, 57–76.
- Jiménez Catalán, R. M., & Ojeda Alba, J. (2009b). Disponibilidad léxica en inglés como lengua extranjera en dos tipos de instrucción. *Lenguaje y Textos*, 30, 167-176.
- Jiménez Catalán, R. M., & Ruiz de Zarobe, Y. (2009). The receptive vocabulary of EFL learners in two instructional contexts: CLIL versus non-CLIL instruction. In R. M. Jiménez Catalán & Y. Ruiz de Zarobe (Eds.), *Content and language integrated learning: Evidence from research in Europe* (pp. 81-92). Bristol: Multilingual Matters.
- Jiménez Catalán, R. M., Agustín Llach, M. P., Fernández Fontecha, A., & Canga Alonso, A. (2014). The effect of age on EFL learners' lexical availability: Word responses to the cue words 'town' and 'countryside'. In R. M. Jiménez Catalán (Ed.), *Lexical Availability in English and Spanish as a Second Language* (pp. 37-51). Dordrecht: Springer.
- Jiménez Catalán, R. M., Ruiz de Zarobe, Y., & Cenoz Iragui, J. (2006). Vocabulary profiles of English foreign language learners in English as a subject and as a vehicular language. *Vienna English Working Papers (Views)*, 15(3), 23-27.
- Jiménez-Catalán, R. M., & Fitzpatrick, T. (2014). Frequency profiles of efl learners' lexical availability. In R. M. Jiménez Catalán (Ed.), *Lexical Availability in English and Spanish as a Second Language* (pp. 83-100). Dordrecht: Springer.
- Johansson, V. (2009). Lexical diversity and lexical density in speech and writing: A developmental perspective. *Lund Working Papers in Linguistics*, 53, 61-79.
- Johnson, W. (1944). Studies in language behavior: A program of research. *Psychological Monographs*, 56(2), 1-15.
- Juan-Garau, M., Prieto-Arranz, J. I., & Salazar-Noguera, J. (2015). Lexico-grammatical development in secondary education CLIL learners. In M. Juan-Garau & J. Salazar-Noguera (Eds.), *Content-based language learning in multilingual educational environments* (pp. 179-195). Cham: Springer.
- Julkunen, K. (1989). *Situation-and task-specific motivation in foreign-language learning and teaching*. Joensuu: University of Joensuu.

- Julkunen, K. (2001). Situation- and task-specific motivation in foreign language learning. In Z. Dörnyei & R. Schmidt (Eds.), *Motivation and Second Language Acquisition* (pp. 29-41). Honolulu, HI: University of Hawaii Press.
- Kamimoto, T. (2005, September). *The effect of guessing on vocabulary test scores: A qualitative analysis*. Paper presented at The European Second Language Association (EuroSLA) Conference 15, Dubrovnik, Croatia.
- Kear, D. J., Coffman, G. A., McKenna, M. C., & Ambrosio, A. L. (2000). Measuring attitude toward writing: A new tool for teachers. *The Reading Teacher, 54*, 10–23.
- Keating, G. D. (2008). Task effectiveness and word learning in a second language: The involvement load hypothesis on trial. *Language Teaching Research, 12*(3), 365-386.
- Keblawi, F. (2009). A review of language learning motivation theories. *Jameea (12)*, 23-57.
- Kelly, P. (1990). Guessing: no substitute for systematic learning of lexis. *System, 18*, 199–208.
- Khatib, M., & Taie, M. (2016). BICS and CALP: Implications for SLA. *Journal of Language Teaching and Research, 7*(2), 382-388.
- Khoshdel-Niyat, F. (2017). The C-Test: A Valid Measure to Test Second Language Proficiency? [Preprint]. HAL Hprints, 01491274, 1–30. <https://doi.org/10.31219/osf.io/c7sy5>
- Kikuchi, K. (2019). Motivation and demotivation over two years A case study of English language learners in Japan. *Studies in Second Language Learning and Teaching, 9*(1), 157-175.
- Kim, T.-Y. (2009). The Dynamics of L2 Self and L2 Learning Motivation: A Qualitative Case Study of Korean ESL Students. *English Teaching, 64*(3), 49–70.
- Kim, T.-Y. (2011). Korean elementary school students' English learning demotivation: A comparative survey study. *Asia Pacific Education Review, 12*(1), 1-11.
- Kimura, D. (1992). Sex differences in the brain. *Scientific American, 267*(3), 118-125.
- Klapper, J. (2008). Deliberate and incidental: vocabulary learning strategies in independent second language learning. In S. Hurd, & T. Lewis (Eds.), *Language*

- Learning strategies in independent settings* (pp. 159-178). Bristol: Multilingual Matters.
- Klein-Braley, C. (1981). *Empirical investigations of cloze tests*. (Unpublished doctoral dissertation). University of Duisburg, Duisberg, Germany.
- Kleinginna, P. R., & Kleinginna, A. M. (1981). A categorized list of motivation definitions, with a suggestion for a consensual definition. *Motivation and Emotion*, 5(3), 263-291.
- Knowles, M. S. (1975). *Self Directed Learning: A Guide for Learners and Teachers*. Chicago: Association Press.
- Kontra, E. H., & Komos, J. (2006). Strategy use and the construct of C-tests. In R. Grotjahn (Ed.), *Der C-Test: Theorie, Empirie, Anwendungen* (pp. 121-138). Frankfurt am Main: Peter Lang.
- Kormos, J. (2011). Task complexity and linguistic and discourse features of narrative writing performance. *Journal of Second Language Writing*, 20(2), 148-161.
- Kormos, J., Csizér, K., & Iwaniec, J. (2014). A mixed-method study of language-learning motivation and intercultural contact of international students. *Journal of Multilingual and Multicultural Development*, 35(2), 151-166.
- Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *The Modern Language Journal*, 73(4), 440-464.
- Krashen, S. D. (1976). Formal and informal linguistic environments in language acquisition and language learning. *TESOL Quarterly*, 10(2), 157-168.
- Krashen, S. D. (1985). *The Input Hypothesis: Issues and Implications*. Harlow: Longman.
- Kremmel, B., & Schmitt, N. (2018). Vocabulary Levels Test. In J. L. Liantas (Ed.), *The TESOL Encyclopedia of English Language Teaching*. John Wiley & Sons. <https://doi.org/10.1002/9781118784235.eelt0499>
- Krzemińska-Adamek M. (2016). Lexis in Writing: Investigating the Relationship Between Lexical Richness and the Quality of Advanced Learners' Texts. In M. Pawlak (Ed.), *Classroom-Oriented Research. Reconciling Theory and Practice* (pp. 185-196). Cham: Springer.

- Lahuerta Martínez, A. C. (2020). Analysis of accuracy in the writing of EFL students enrolled on CLIL and non-CLIL programmes: the impact of grade and gender. *The Language Learning Journal*, 48(2), 121-132.
- Lamb, M. (2007). The impact of school on EFL learning motivation: An Indonesian case study. *TESOL Quarterly*, 41(4), 757-780.
- Lamb, M. (2012). A self system perspective on young adolescents' motivation to learn English in urban and rural settings. *Language Learning*, 62(4), 997-1023.
- Lamb, M. (2017). The motivational dimension of language teaching. *Language Teaching*, 50(3), 301-346.
- Larson-Hall, J. (2012). How to Run Statistical Analyses. In A. Mackey & S. M. Gass (Eds.), *Research Methods in Second Language Acquisition: A Practical Guide* (pp. 245-274). Oxford: Wiley-Blackwell.
- Larson-Hall, J. (2015). *A guide to doing statistics in second language research using SPSS and R*. New York: Routledge.
- Larsson, J. (2001). *Problem-Based Learning: A possible approach to language education*. Retrieved from <https://www.nada.kth.se/~jla/docs/PBL.pdf>
- Lasagabaster, D. (2008). Foreign language competence in content and language integrated courses. *The Open Applied Linguistics Journal*, 1, 30-41.
- Lasagabaster, D. (2009). The implementation of CLIL and attitudes towards trilingualism. *ITL-International Journal of Applied Linguistics*, 157(1), 23-43.
- Lasagabaster, D. (2011). English achievement and student motivation in CLIL and EFL settings. *Innovation in language Learning and Teaching*, 5(1), 3-18.
- Lasagabaster, D. (2014). The impact of gender and English proficiency on attitudes towards English-medium instruction. *Rivista di Psicolinguistica Applicata/Journal of Applied Psycholinguistics*, 14, 67-82.
- Lasagabaster, D. (2016). The relationship between motivation, gender, L1 and possible selves in English-medium instruction. *International Journal of Multilingualism*, 13(3), 315-332.
- Lasagabaster, D. (2019). Motivation in Content and Language Integrated Learning (CLIL) Research. In M. Lamb, K. Csizér, A. Henry, & S. Ryan (Eds.), *The Palgrave*

- Handbook of Motivation for Language Learning* (pp. 347-366). Cham: Palgrave Macmillan.
- Lasagabaster, D., & Doiz, A. (2017). A longitudinal study on the impact of CLIL on affective factors. *Applied Linguistics*, 38(5), 688-712.
- Lasagabaster, D., & López Belouqui, R. (2015). The impact of type of approach (CLIL versus EFL) and methodology (book-based versus project work) on motivation. *Porta Linguarum: Revista Internacional de Didáctica de las Lenguas Extranjeras*, 23, 41-57.
- Lasagabaster, D., & Ruiz de Zarobe, Y. (2010). *CLIL in Spain: Implementation, results and teacher training*. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Lasagabaster, D., & Sierra, J. M. (2009). Language attitudes in CLIL and traditional EFL classes. *International CLIL Research Journal*, 1(2), 4-17.
- Lasagabaster, D., Doiz, A., & Sierra, J. M. (2014). Introduction. In D. Lasagabaster, A. Doiz & J. M. Sierra (Eds.), *Motivation and foreign language learning: From theory to practice* (pp. 1-5). Amsterdam: John Benjamins Publishing Company.
- Laufer, B. (1998). The development of passive and active vocabulary in a second language: same or different? *Applied Linguistics*, 19(2), 255-271.
- Laufer, B. (2009). Second language vocabulary acquisition from language input and from form-focused activities. *Language Teaching*, 42(3), 341-354.
- Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, 54(3), 399-436.
- Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, 22(1), 1-26.
- Laufer, B., & Nation, I. S. P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, 16(3), 307-322.
- Laufer, B., & Nation, I. S. P. (2012). Vocabulary. In S. M. Gass & A. Mackey (Eds.), *The Routledge Handbook of Second Language Acquisition* (pp. 163– 176). New York: Routledge.
- Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing*, 16(1), 33-51.

- Laufer, B., & Nation, P. (2001). Passive vocabulary size and speed of meaning recognition: Are they related? *Eurosla Yearbook*, 1(1), 7-28.
- Laufer, B., & Paribakht, T. S. (1998). The relationship between passive and active vocabularies: Effects of language learning context. *Language Learning*, 48(3), 365-391.
- Leeper, C., & Smith, T.E. (2004). A meta-analytic review of gender variations in children's language use: 1012 Talkativeness, affiliative speech, and assertive speech. *Developmental Psychology*, 40(6), 993-1027.
- Lee, J. (2020). Assessing the effects of CLIL on Korean high school students' writing. *Linguistic Research*, 37, 89-112.
- Lee, S., & Pulido, D. (2017). The impact of topic interest, L2 proficiency, and gender on EFL incidental vocabulary acquisition through reading. *Language Teaching Research*, 21(1), 118-135.
- Leńko-Szymańska, A. (2019). *Defining and Assessing Lexical Proficiency*. New York: Routledge.
- Lessard-Clouston, M. (2013). *Teaching Vocabulary*. Alexandria, VA: TESOL International Association.
- Lew, S., Yang, A. H., & Harklau, L. (2018). Qualitative methodology. In A. Phakiti, P. de Costa, L. Plonsky & S. Starfield (Eds.), *The Palgrave handbook of applied linguistics research methodology* (pp. 79-101). London: Palgrave Macmillan
- Li, Z., Chen, M. Y., & Banerjee, J. (2020). Using corpus analyses to help address the DIF interpretation: Gender differences in standardized writing assessment. *Frontiers in Psychology*, 11, 1088. <https://doi.org/10.3389/fpsyg.2020.01088>
- Lightbown, P. M. (2014). *Focus on content-based language teaching*. Oxford: Oxford University Press.
- Lindqvist, C. (2020). First and second language cognate effects in third language vocabulary size estimates. In P. Booth and J. Clenton (Eds), *First Language Influences on Multilingual Lexicons* (pp. 149-161). New York: Routledge.

- Lindqvist, C., Bardel, C., & Gudmundson, A. (2011). Lexical richness in the advanced learner's oral production of French and Italian L2. *International Review of Applied Linguistics in Language Teaching*, 49(3), 221-240.
- Lindqvist, C., Gudmundson, A., & Bardel, C. (2013). A new approach to measuring lexical sophistication in L2 oral production. In C. Bardel, C. Lindqvist & B. Laufer (Eds.), *L2 vocabulary acquisition, knowledge and use: New perspectives on assessment and corpus analysis* (pp. 109-126). Eurosla Monographs Series 2. Amsterdam: Eurosla.
- Linnarud, M. (1976). Lexical density and lexical variation-An analysis of the lexical texture of Swedish students' written work. *Studia Anglica Posnaniensia: An International Review of English Studies*, 7, 45-52.
- Littlewood, W. (1996). "Autonomy": An anatomy and a framework. *System*, 24(4), 427-435.
- Liyanage, I., & Bartlett, B. J. (2012). Gender and language learning strategies: Looking beyond the categories. *The Language Learning Journal*, 40(2), 237-253.
- Lo, Y., & Macaro, E. (2015). Getting used to content and language integrated learning: What can classroom interaction reveal? *The Language Learning Journal*, 43(3), 1-17.
- Long, M. (1991). Focus on Form: A design feature in language teaching methodology. In K. de Bot, R. Ginsberg, & C. Kramsch (Eds.), *Foreign language research in cross-cultural perspective* (pp. 39-52). Amsterdam: John Benjamins.
- Long, M. H. (1981). Input, interaction, and second language acquisition. *Annals of the New York Academy of Sciences*, 379(1), 259-278.
- Long, M. H. and Richards, J. C. (2007). Series editors' preface. In H. Daller, J. Milton and J. Treffers-Daller (Eds.), *Modelling and assessing vocabulary knowledge* (pp. xii-xiii). Cambridge, UK: Cambridge University Press.
- Lonsdale, D., & Le Bras, Y. (2009). *A frequency dictionary of French: Core vocabulary for learners*. London: Routledge.
- López Morales, H. (1973). Disponibilidad léxica de los escolares de San Juan. *San Juan de Puerto Rico: Universidad de Puerto Rico*.

- López Morales, H. L. (2014). Lexical availability studies. In R. M. Jiménez Catalán (Ed.), *Lexical Availability in English and Spanish as a Second Language* (pp. 1-11). Dordrecht: Springer.
- López Rúa, P. (2006). The sex variable in foreign language learning an integrative approach. *Porta Linguarum: Revista Internacional de Didáctica de las Lenguas Extranjeras*, 6, 99-114.
- Lorenzo, F. (2007). The sociolinguistics of CLIL: Language planning and language change in 21st century Europe. *Revista Española de Lingüística Aplicada*, 20(1), 27-38.
- Lorenzo, F., Casal, S., & Moore, P. (2010). The effects of content and language integrated learning in European education: Key findings from the Andalusian bilingual sections evaluation project. *Applied Linguistics*, 31(3), 418-442.
- Lu, M. (2013). Effects of four vocabulary exercises on facilitating learning vocabulary meaning, form, and use. *TESOL Quarterly*, 47(1), 167-176.
- Lu, Y., Bao, G., & Cui, B. L. (2013). A Comparison of Three Measures of Lexical Sophistication for EFL Learners. *Journal of PLA University of Foreign Languages*, 5, 44-48.
- Ludwig, J. (1983). Attitudes and expectations: A profile of female and male students of college French, German and Spanish. *The Modern Language Journal*, 67(3), 216–227.
- Lumsden, K. G., & Scott, A. (1987). The economics student reexamined: Male-female differences in comprehension. *The Journal of Economic Education*, 18(4), 365-375.
- Lyster, R., & Ballinger, S. (2011). Content-based language teaching: Convergent concerns across divergent contexts. *Language teaching research*, 15(3), 279-288.
- Ma, Q. (2009). *Second language vocabulary acquisition*. Bern, Switzerland: Peter Lang.
- Maas, H.D. (1972). Zusammenhang zwischen Wortschatzumfang und Länge eines Textes. *Zeitschrift für Literaturwissenschaft und Linguistik*, 8, 73–79.
- MacIntyre, P. D., & Serroul, A. (2015). Motivation on a per-second timescale: Examining approach-avoidance motivation during L2 task performance. In Z. Dörnyei, P. D.

- MacIntyre, & A. Henry (Eds.), *Motivational Dynamics in Language Learning* (pp. 109-138). Bristol, UK: Multilingual Matters.
- MacIntyre, P. D., MacKinnon, S. P., & Clément, R. (2009). Toward the development of a scale to assess possible selves as a source of language learning motivation. In Z. Dörnyei, & E. Ushioda (Eds.), *Motivation, Language Identity and the L2 Self* (pp. 193-214). Bristol: Multilingual Matters.
- Mackey, W. F. (1971). *Le vocabulaire disponible du français*. Paris: Didier.
- Macksoud, R. (2010). Using Interview Data in Case Studies. In S. Hunston & D. Oakley (Eds.), *Introducing applied linguistics: Concepts and skills* (pp. 153–159). London: Routledge.
- Madrid Fernández, D., Ortega-Martín, J. L., & Hughes, S. P. (2019). CLIL and Language Education in Spain. In K. Tsuchiya & D. Pérez Murillo (Eds.), *Content and Language Integrated Learning in Spanish and Japanese Contexts* (pp. 11-35). Cham: Palgrave Macmillan.
- Malvern, D. D., & Richards, B. J. (1997). A new measure of lexical diversity. *British Studies in Applied Linguistics*, 12, 58-71.
- Malvern, D., & Richards, B. (2002). Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language Testing*, 19(1), 85-104.
- Manzano Vázquez, B. (2015). CLIL in three Spanish monolingual communities: the examples of Extremadura, Madrid and La Rioja. *Elia: Estudios de Lingüística Inglesa Aplicada*, 15, 135-158.
- Markus, H., & Nurius, P. (1986). Possible selves. *American Psychologist*, 41(9), 954.
- Marsh, D. (2000). *Using languages to learn and learning to use languages*. Finland: University of Jyväskylä.
- Marsh, D. (2002). *CLIL-EMILE The European Dimension: Actions, Trends and Foresight Potential*. Brussels: European Commission.
- Marsh, D. (2009). Forward. In Y. Ruiz de Zarobe & R. M. Jiménez Catalán (Eds.), *Content and language integrated learning Evidence from research in Europe* (pp. vii-viii). Bristol: Multilingual Matters.

- Martinez-Adrian, M., & Gallardo-del-Puerto, F. (2017). The effects of language typology on L2 lexical availability and spelling accuracy. *International Journal of English Studies, 17*(2), 63-79.
- Martinsen, R. A., Alvord, S. M., & Tanner, J. (2014). Perceived foreign accent: Extended stays abroad, level of instruction, and motivation. *Foreign Language Annals, 47*(1), 66-78.
- Matiasek, S. (2005). *English as the language of instruction in Austrian chemistry lessons: Instances of explicit language teaching*. (Unpublished master's thesis). University of Vienna, Austria.
- McCarthy, D. (1954). Language development in children. In L. Carmichael (Ed.), *Manual of child psychology* (2nd ed., pp. 491-630). New York: Wiley.
- McCarthy, P. M. (2005). *An assessment of the range and usefulness of lexical diversity measures and the potential of the measure of textual, lexical diversity (MTLD)* (Unpublished doctoral dissertation). The University of Memphis, Memphis, Tennessee.
- McCarthy, P. M., & Jarvis, S. (2007). vocd: A theoretical and empirical evaluation. *Language Testing, 24*(4), 459-488.
- McEown, K., & Sugita-McEown, M. (2020). The role of positive and negative psychological factors in predicting effort and anxiety toward languages other than English. *Journal of Multilingual and Multicultural Development*. <https://doi.org/10.1080/01434632.2020.1767634>
- McKenna, M. C., & Kear, D. J. (1999). Measuring attitude toward reading: A new tool for teachers. In S. J. Barrentine (Ed.), *Reading assessment. Principles and practices for elementary teachers* (pp. 199–214). Newark, DE: International Reading Association.
- McLean, S., & Kramer, B. (2015). The creation of a new vocabulary levels test. *Shiken, 19*(2), 1–11.
- McLean, S., & Kramer, B. (2016). The development of a Japanese bilingual version of the New Vocabulary Levels Test. *Vocabulary Education and Research Bulletin, 5*(1), 2-5.

- McLean, S., Kramer, B., & Beglar, D. (2015). The creation and validation of a listening vocabulary levels test. *Language Teaching Research*, 19(6), 741-760.
- McManus, K. (2011). *The development of aspect in a second language* (Unpublished doctoral dissertation). Newcastle University, Newcastle, England.
- Meara, P. (1996). The dimensions of lexical competence. In G. Brown, K. Malmkjaer and J. Williams (Eds.), *Performance and Competence in Second Language Acquisition*. (pp. 35- 53). Cambridge, UK: Cambridge University Press.
- Meara, P. (1997). Towards a new approach to modelling vocabulary acquisition. In N. Schmitt & M. Mc Carthy (Eds.), *Vocabulary: Description, Acquisition and Pedagogy* (pp. 109-121). Cambridge. Cambridge University Press.
- Meara, P. (2005). Lexical frequency profiles: A Monte Carlo analysis. *Applied Linguistics*, 26(1), 32-47.
- Meara, P. M., & Milton, J. (2003). *X-lex: the Swansea levels test*. Newbury: Express Publishing.
- Meara, P. M., & Miralpeix, I. (2006). Y_Lex: The Swansea advanced vocabulary levels test. v2. 05. *Swansea: Lognostics*.
- Meara, P., & Bell, H. (2001). P_Lex: A simple and effective way of describing the lexical characteristics of short L2 texts. *Prospect*, 16(3), 323–37.
- Meara, P., & Buxton, B. (1987). An alternative to multiple choice vocabulary tests. *Language Testing*, 4(2), 142-154.
- Meara, P., & Fitzpatrick, T. (2000). Lex30: An improved method of assessing productive vocabulary in an L2. *System*, 28(1), 19-30.
- Meara, P., & Jones, G. (1990). *The Eurocentres Vocabulary Size Test. 10K*. Zurich: Eurocentres.
- Meara, P., & Miralpeix, I. (2004). V_Size. *Swansea: Lognostics*.
- Meara, P., & Miralpeix, I. (2015). V_YesNo v1 [Computer Program]. Retrieved from http://www.lognostics.co.uk/tools/V_YesNo/V_YesNo.htm
- Meara, P., & Miralpeix, I. (2016). *Tools for researching vocabulary*. Bristol, UK: Multilingual Matters.

- Mehisto, P., Marsh, D., & Frigols, M. J. (2008). *Uncovering CLIL content and language integrated learning in bilingual and multilingual education*. Oxford: Macmillan.
- Mehrpour, S., & Vojdani, M. (2012). Globalization and EFL Learning Motivation: A New Perspective on Integrative vs. Instrumental Motivation among Iranian Learners of English. *Open Journal of Modern Linguistics*, 2(2), 43–50.
- Melka Teichroew, F. J. (1982). Receptive versus productive vocabulary: A survey. *Interlanguage Studies Bulletin*, 6(2), 5-33.
- Melka, F. (1997). Receptive vs. productive aspects of vocabulary. In N. Schmitt and M. McCarthy (Eds), *Vocabulary: Description, Acquisition and Pedagogy* (pp. 84–102). Cambridge, UK: Cambridge University Press.
- Mendoza, A., & Phung, H. (2019). Motivation to learn languages other than English: A critical research synthesis. *Foreign Language Annals*, 52(1), 121-140.
- Merikivi, R., & Pietilä, P. (2014). Vocabulary in CLIL and in mainstream education. *Journal of Language Teaching and Research*, 5(3), 487.
- Merisuo-Storm, T. (2006). Girls and boys like to read and write different texts. *Scandinavian Journal of educational research*, 50(2), 111-125.
- Merisuo-Storm, T. (2007). Pupils' attitudes towards foreign-language learning and the development of literacy skills in bilingual education. *Teaching and Teacher Education*, 23(2), 226-235.
- Michéa, R. (1953). Mots fréquents et mots disponibles. Un aspect nouveau de la statistique du langage. *Les Langues Modernes*, 47, 338–344.
- Milton, J. (2009). *Measuring second language vocabulary acquisition*. Clevedon, UK: Multilingual Matters.
- Milton, J., & Donzelli, G. (2013). The Lexicon. In J. R. Herschensohn & M. Young-Scholten (Eds.), *The Cambridge handbook of second language acquisition* (pp. 441-460). Cambridge, UK: Cambridge University Press.
- Milton, J., & Hopkins, N. (2006). Comparing phonological and orthographic vocabulary size: Do vocabulary tests underestimate the knowledge of some learners? *Canadian Modern Language Review*, 63(1), 127-147.

- Miralpeix, I. (2008). *The Influence of Age on Vocabulary Acquisition in English as a Foreign Language*. (Unpublished doctoral dissertation). University of Barcelona, Spain.
- Mochida, K., & Harrington, M. (2006). The Yes/No test as a measure of receptive vocabulary knowledge. *Language Testing*, 23(1), 73-98.
- Montero-Saiz Aja, A. (2021). Gender-based Differences in EFL Learners' Language Learning Strategies and Productive Vocabulary. *Theory and Practice of Second Language Acquisition*, 7(2), 83-107.
- Mora, J. C., & Valls-Ferrer, M. (2012). Oral fluency, accuracy, and complexity in formal instruction and study abroad learning contexts. *TESOL Quarterly*, 46(4), 610-641.
- Moreno Espinosa, S. (2003). Vocabulary: Reviewing Trends in EFL/ESL Instruction and Testing. *Odisea*, 4, 91-112.
- Moreno Espinosa, S. (2009). Young learners' L2 word association responses in two different learning contexts. In R. M. Jiménez Catalán & Y. Ruiz de Zarobe (Eds.), *Content and language integrated learning: Evidence from research in Europe* (pp. 93-111). Bristol: Multilingual Matters.
- Moreno Espinosa, S. (2010). Boys' and girls' L2 word associations. In R. M. Jiménez Catalán (Ed.), *Gender Perspectives on Vocabulary in Foreign and Second languages* (pp. 139-163). New York: Palgrave Macmillan.
- Moreno Espinosa, S., & Agustín Llach, M. P. (2010). What Electronic Tools Can EFL Teachers Use to Assess L2 Embedded Vocabulary Objectively? *Malaysian Journal of ELT Research*, 6(1), 86-132.
- Morgan, B. Q., & Oberdeck, L. M. (1930). Active and passive vocabulary. In E. W. Bagster-Collins (Ed.), *Studies in Modern Language Teaching* (pp. 213-221). London: Macmillan.
- Mori, S., & Gobel, P. (2006). Motivation and gender in the Japanese EFL classroom. *System*, 34(2), 194-210.
- Moskovsky, C., Assulaimani, T., Racheva, S., & Harkins, J. (2016). The L2 motivational self system and L2 achievement: A study of Saudi EFL learners. *The Modern Language Journal*, 100(3), 641-654.

- Muncie, J. (2002). Process writing and vocabulary development: Comparing lexical frequency profiles across drafts. *System*, 30(2), 225-235.
- Murayama, K. (2018). The Science of Motivation. Multidisciplinary Approaches Advance Research on the Nature and Effects of Motivation. *Psychological Science Agenda*, June 2018.
- Nation, I. (2006). How large a vocabulary is needed for reading and listening. *Canadian Modern Language Review*, 63(1), 59-82.
- Nation, I. S. P. (1982). Beginning to learn foreign vocabulary: A review of the research. *RELC Journal*, 13(1), 14-36.
- Nation, I. S. P. (1983). Testing and teaching vocabulary. *Guidelines*, 5, 12-25.
- Nation, I. S. P. (1984). *Vocabulary lists: Words, affixes and stems*. English University of Wellington, New Zealand: English Language Institute.
- Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Boston, MA: Heinle and Heinle.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge, UK: Cambridge University Press.
- Nation, I. S. P. (2004). A study of the most frequent word families in the British National Corpus. *Vocabulary in a second language: Selection, Acquisition, and Testing*, 10, 3-13.
- Nation, I. S. P. (2006). Language education - vocabulary. In K. Brown (Ed.), *Encyclopaedia of Language and Linguistics* (2nd ed., pp. 494-499). Oxford: Elsevier.
- Nation, I. S. P. (2012). *The BNC/COCA word family lists*. Retrieved from: <http://www.victoria.ac.nz/lals/about/staff/paul-nation>
- Nation, I. S. P. (2016). *Making and using word lists for language learning and testing*. Amsterdam: John Benjamins Publishing Company.
- Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, 31(7), 9-13.

- Nation, I. S. P., & Chung, T. (2011). Teaching and Testing Vocabulary. In M. H. Long, & C. J. Doughty (Eds.), *The Handbook of Language Teaching* (pp. 543-559). Malden, MA: John Wiley & Sons.
- Nation, I. S. P., & Meara, P. (2013). Vocabulary. In N. Schmitt, (Ed.), *An Introduction to Applied Linguistics* (2nd ed., pp. 34-52). London: Routledge.
- Nation, I. S. P., & Waring, R. (1997). Vocabulary size, text coverage and word lists. *Vocabulary: Description, Acquisition and Pedagogy*, 14, 6-19.
- Navarro Pablo, M. (2018). Are CLIL students more motivated an analysis of affective factors and their relation to language attainment. *Porta Linguarum: Revista Internacional de Didáctica de las Lenguas Extranjeras*, 29, 71-90.
- Navés, T. (2011). How Promising are the Results of Integrating Content and Language for EFL Writing and Overall EFL Proficiency? In Y. Ruiz de Zarobe, J. M. Sierra & F. Gallardo del Puerto (Eds), *Content and Foreign Language Integrated Learning: Contributions to Multilingualism in European Contexts* (pp. 155-186). Bern: Peter Lang.
- Navés, T., & Victori, M. (2010). CLIL in Catalonia: An overview of research studies. In D. Lasagabaster & Y. Ruiz de Zarobe (Eds), *CLIL in Spain: Implementation, results and teacher training* (pp. 30-54). Newcastle upon Tyne: Cambridge Scholars Publishing.
- Neilson Parada, M. (2016). *Lexical Availability in Diaspora Spanish: A Cross-generational Analysis of Chilean Swedes*. (Unpublished doctoral dissertation). University of Illinois at Chicago, Chicago, Illinois.
- Niemiec, C. P., & Ryan, R. M. (2009). Autonomy, competence, and relatedness in the classroom: Applying self-determination theory to educational practice. *Theory and Research in Education*, 7(2), 133-144.
- Nikula T. (2017). CLIL: A European Approach to Bilingual Education. In N. van Deusen-Scholl & S. May (Eds.), *Second and Foreign Language Education. Encyclopedia of Language and Education* (3rd ed., pp. 111-124). Cham: Springer.

- Nikula, T. (2010). Effects of CLIL on a teacher's classroom language use. In C. Dalton-Puffer, T. Nikula, & U. Smit (Eds.), *Language use and language learning in CLIL classrooms* (pp. 105-124). Amsterdam: John Benjamins.
- Noels, K. A. (2001). Learning Spanish as a second language: Learners' orientations and perceptions of their teachers' communication style. *Language Learning*, 51(1), 107-144.
- Noels, K. A., Lou, N. M., Vargas Lascano, D. I., Chaffee, K. E., Dincer, A., Zhang, Y. S. D. & Zhang, X. (2019). Self-determination and motivated engagement in language learning. In M. Lamb, K. Csizér, A. Henry & S. Ryan (Eds.), *The Palgrave Handbook of Motivation for Language Learning* (pp. 95-115). Cham: Palgrave Macmillan.
- Noels, K. A., Pelletier, L. G., Clément, R., & Vallerand, R. J. (2000). Why are you learning a second language? Motivational orientations and self-determination theory. *Language Learning*, 50(1), 57-85.
- Norris, J. M. (2018). *Developing C-tests for estimating proficiency in foreign language research*. Frankfurt am Main, Germany: Peter Lang.
- Nurweni, A., & Read, J. (1999). The English vocabulary knowledge of Indonesian university students. *English for Specific Purposes*, 18(2), 161-175.
- Oakes, L. (2013). Foreign language learning in a 'monoglot culture': Motivational variables amongst students of French and Spanish at an English university. *System*, 41(1), 178-191.
- Oakes, L., & Howard, M. (2019). Learning French as a foreign language in a globalised world: an empirical critique of the L2 Motivational Self System. *International Journal of Bilingual Education and Bilingualism*, 25(1), 166-182.
- Oakley, A. (1972). *Sex, gender and Society*. London: Temple Smith.
- Ogden, C. K. (1930). *Basic English: A General Introduction*. London: Kegan Paul, Trench & Trubner.
- Ojeda Alba, J. (2009). Themes and vocabulary in CLIL and non-CLIL instruction. In R. M. Jiménez Catalán & Y. Ruiz de Zarobe (Eds.), *Content and language integrated learning: Evidence from research in Europe* (pp. 130-156). Bristol: Multilingual Matters.

- Olsson, E. (2015). Progress in English academic vocabulary use in writing among CLIL and non-CLIL students in Sweden. *Moderna Språk, 109*(2), 51-74.
- Ortega, L. (2014). *Understanding second language acquisition*. London: Routledge.
- Ortega, L., & Byrnes, H. (2008). The longitudinal study of advanced L2 capacities: An introduction. In L. Ortega & H. Byrnes (Eds.), *The Longitudinal Study of Advanced L2 Capacities* (pp. 3–20). New York: Routledge.
- Ortega, L., & Ibarra-Shea, G. (2005). Longitudinal research in second language acquisition: Recent trends and future directions. *Annual Review of Applied Linguistics, 25*, 26–45.
- Ortega-Martín, J. L., & Trujillo, F. (2018). Legislación y normativa para el funcionamiento de los programas AICLE en España. In J. L. Ortega-Martín, S. P. Hughes & D. Madrid (Eds.), *Influencia de la política educativa en la enseñanza bilingüe* (pp. 21-30). Madrid: MECD-British Council, Secretaría General Técnica, Ministerio de Educación, Cultura y Deporte.
- Ortega-Martín, J. L., Hughes, S. P., & Madrid, D. (Eds.) (2018). *Influencia de la política educativa de centro en la enseñanza bilingüe en España*. Madrid: MECD-British Council, Secretaría General Técnica, Ministerio de Educación, Cultura y Deporte.
- Otwinowska, A., & Foris, M. (2017). They learn the CLIL way, but do they like it? Affectivity and cognition in upper-primary classes. *International Journal of Bilingual Education and Bilingualism, 20*(5), 457–480.
- Ovtcharov, V., Cobb, T., & Halter, R. (2006). La richesse lexicale des productions orales: mesure fiable du niveau de compétence langagière. *Canadian Modern Language Review, 63*(1), 107-125.
- Oxford, R. (1998, March). *The unravelling tapestry: Teacher and course characteristics associated with demotivation in the language classroom: Demotivation in foreign language learning*. Paper presented at the TESOL '98 Congress, Seattle, WA.
- Oxford, R. L. (1989). Use of language learning strategies: A synthesis of studies with implications for strategy training. *System, 17*(2), 235-247.
- Oxford, R. L. (1994). Where are we with language learning motivation? *The Modern Language Journal, 78*(4), 512–514.

- Oxford, R. L., & Shearin, J. (1994). Language learning motivation: Expanding the theoretical framework. *The Modern Language Journal*, 78(1), 12–28.
- Oxford, R. L., Nyikos, M., & Ehrman, M. (1988). Vive la Difference? Reflections on Sex Differences in Use of Language Learning Strategies. *Foreign Language Annals*, 21(4), 321–329.
- Pae, T. I. (2004). Gender effect on reading comprehension with Korean EFL learners. *System*, 32(2), 265-281.
- Palmberg, R. (1987). Patterns of vocabulary development in foreign-language learners. *Studies in Second Language Acquisition*, 9(2), 201-219.
- Palmer, H. E. (1921). *The Principles of Language Study*. London: Harrap.
- Papi, M., & Abdollahzadeh, E. (2012). Teacher motivational practice, student motivation, and possible L2 selves: An examination in the Iranian EFL context. *Language Learning*, 62(2), 571-594.
- Papi, M., Bondarenko, A. V., Mansouri, S., Feng, L., & Jiang, C. (2019). Rethinking L2 motivation research: The 2× 2 model of L2 self-guides. *Studies in Second Language Acquisition*, 41(2), 337-361.
- Paran, A. (2013). Content and language integrated learning: Panacea or policy borrowing myth? *Applied Linguistics Review*, 4(2), 317-342.
- Pavesi, M., Bertocchi, D., Hofmannová, M., & Kazianka, M. (2001). Teaching through a Foreign Language: A guide for teachers and schools to using foreign languages in content teaching. In G. Langé (Ed.), *CLIL Guidelines for Teachers* (pp. 72-102). Milan: TIE CLIL.
- Pavón Vázquez, V., & Rubio, F. (2010). Teachers' concerns and uncertainties about the introduction of CLIL programmes. *Porta Linguarum: Revista Internacional de Didáctica de las Lenguas Extranjeras*, 14, 45-58.
- Payne, C. (2016). Lexical availability. *English Teaching Professional*, 102, 18-20.
- Pelechano, V. (1994). Prueba MA. *Análisis y Modificación de la Conducta*, 20, 71-72.
- Pérez Cañado, M. L. (2011). The effects of CLIL within the APPP: lessons learned and ways forward. In R. Crespo & M. García de Sola (Eds.), *Studies in honour of Ángeles Linde López* (pp. 389-406). Granada: Universidad de Granada.

- Pérez Cañado, M. L. (2018a). CLIL and educational level: A longitudinal study on the impact of CLIL on language outcomes. *Porta Linguarum: Revista Internacional de Didáctica de las Lenguas Extranjeras*, 29, 51-70.
- Pérez Cañado, M. L. (2018b). The effects of CLIL on L1 and content learning: Updated empirical evidence from monolingual contexts. *Learning and Instruction*, 57, 18-33.
- Pérez Cañado, M. L., & Lancaster, N. K. (2017). The effects of CLIL on oral comprehension and production: A longitudinal case study. *Language, Culture and Curriculum*, 30(3), 300-316.
- Pérez Cañado, M. L., & Ráez Padilla, J. (2015). Introduction and Overview. In D. Marsh, M. L. Pérez Cañado & J. Ráez Padilla (Eds.), *CLIL in Action: Voices from the Classroom* (pp. 1-12). Newcastle upon Tyne: Cambridge Scholars Publishing.
- Pérez, A., Lorenzo, F., & Pavón, V. (2016). European bilingual models beyond linguafranca: key findings from CLIL French programs. *Language policy*, 15(4), 485-504.
- Pérez-Cañado, M. L. (2012). CLIL research in Europe: Past, present, and future. *International Journal of Bilingual Education and Bilingualism*, 15(3), 315-341.
- Pérez-Vidal, C. (2007). The need for focus on form (FoF) in content and language integrated approaches: An exploratory study. *Revista Española De Lingüística Aplicada*, 1, 39-54.
- Pérez-Vidal, C. (2014) Study abroad and formal instruction contrasted. In C. Pérez-Vidal (Ed.), *Language Acquisition in Study Abroad and Formal Instruction Contexts* (pp. 17–58). Amsterdam: John Benjamins.
- Perry, E. (1998). A quantitative approach to evaluating L2 written English. *Cahiers de l'APLIUT*, 18(1), 66-81.
- Pfenninger, S. (2016). All good things come in threes Early English learning, CLIL and motivation in Switzerland. *Cahiers de IILSL*, 48, 119-147.
- Pignot-Shahov, V. (2012). Measuring L2 receptive and productive vocabulary knowledge. *Language Studies Working Papers*, 4(1), 37-45.

- Piller, I., & Pavlenko, A. (2001). Introduction: Multilingualism, second language learning, and gender. In A. Pavlenko, A. Blackledge, I. Piller & M. Teutsch Dwyer, M (Eds.). *Multilingualism, second language learning, and gender*. Berlin: Mouton de Gruyter.
- Pladevall-Ballester, E. (2015). Exploring primary school CLIL perceptions in Catalonia: students', teachers' and parents' opinions and expectations. *International Journal of Bilingual Education and Bilingualism*, 18(1), 45-59.
- Pladevall-Ballester, E. (2019). A longitudinal study of primary school EFL learning motivation in CLIL and non-CLIL settings. *Language Teaching Research*, 23(6), 765-786.
- Politzer, R. (1983). An exploratory study of self-reported language learning behaviors and their relation to achievement. *Studies in Second Language Acquisition*, 6, 54–67.
- Portocarrero, J. S., Burrig, R. G., & Donovan, P. J. (2007). Vocabulary and verbal fluency of bilingual and monolingual college students. *Archives of Clinical Neuropsychology*, 22(3), 415-422.
- Poupore, G. (2016). Measuring group work dynamics and its relation with L2 learners' task motivation and language production. *Language Teaching Research*, 20(6), 719-740.
- Powell, R., & Littlewood, P. (1983). Why Choose French? Boys' and Girls' Attitudes at the Option Stage. *British Journal of Language Teaching*, 21(1), 36.
- Prieto-Arranz, J. I., Fabra, L. R., Calafat-Ripoll, C., & Catrain-González, M. (2015). Testing progress on receptive skills in CLIL and non-CLIL contexts. In *Content-based language learning in multilingual educational environments* (pp. 123-137). Cham: Springer.
- Qian, D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *Canadian Modern Language Review*, 56(2), 282-308.
- Qian, D. (2000). *Validating the role of depth of vocabulary knowledge in assessing reading for basic comprehension* [TOEFL 2000 Research Report]. Princeton, NJ: Educational Testing Service.

- Raatz, U., & Klein-Braley, C. (2002). Introduction to language testing and to C-Tests. In A. Coleman, R. Grotjahn & U. Raatz (Eds.), *University language testing and the C-Test* (pp. 75- 87). Bochum: AKS-Verlag.
- Ráez Padilla, J. (2018). Parent perspectives on CLIL implementation: Which variables make a difference? *Porta Linguarum: Revista Internacional de Didáctica de las Lenguas Extranjeras*, 29, 181-196.
- Raven, J. (1989). The Raven Progressive Matrices: A review of national norming studies and ethnic and socioeconomic variation within the United States. *Journal of Educational Measurement*, 26, 1-16.
- Read, J. (1988). Measuring the vocabulary knowledge of second language learners. *RELC Journal*, 19, 12–25.
- Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing*, 10(3), 355-371.
- Read, J. (1998). Validating a Test to Measure Depth of Vocabulary Knowledge. In A. Kunnan (Ed.), *Validation in Language Assessment* (pp. 41-60). Mahwah, NJ: Erlbaum.
- Read, J. (2000). *Assessing vocabulary*. Cambridge, UK: Cambridge University Press.
- Read, J. (2004). Plumbing the depths: How should the construct of vocabulary knowledge be defined? In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language* (pp. 209– 227). Amsterdam: John Benjamins.
- Richards, B. (1987). Type/token ratios: What do they really tell us? *Journal of child language*, 14(2), 201-209.
- Richards, J. C. (1976). The role of vocabulary teaching. *TESOL Quarterly*, 10, 77-89.
- Rieder-Bünemann, A., Hüttner, J., & Smit, U. (2022). ‘Who would have thought that I’d ever know that!’: subject-specific vocabulary in CLIL student interactions. *International Journal of Bilingual Education and Bilingualism*. <https://doi.org/10.1080/13670050.2021.2020211>
- Rodriguez Bonces, J. (2012). Content and language integrated learning (CLIL): Considerations in the Colombian context. *GIST: Education and Learning Research Journal*, 6, 177-189.

- Roghani, S., & Milton, J. (2017). Using Category Generation Tasks to Estimate Productive Vocabulary Size in a Foreign Language. *TESOL International Journal*, 12(1), 143-159.
- Ruesch, A., Bown, J., & Dewey, D. P. (2012). Student and teacher perceptions of motivational strategies in the foreign language classroom. *Innovation in Language Learning and Teaching*, 6(1), 15-27.
- Ruiz de Zarobe, Y. (2010). Written production and CLIL: An empirical study. In C. Dalton-Puffer, T. Nikula, and U. Smit (Eds), *Language use and language learning in CLIL classrooms* (pp. 191-209). Amsterdam: John Benjamins.
- Ruiz de Zarobe, Y. (2011). Which language competencies benefit from CLIL? An insight into applied linguistics research. In Y. Ruiz de Zarobe, J. M. Sierra & F. Gallardo del Puerto (Eds.), *Content and Foreign Language Integrated Learning: Contributions to Multilingualism in European Contexts* (pp. 129-154). Bern: Peter Lang.
- Ruiz de Zarobe, Y., & Celaya, M. L. (2011). AICLE a través de las lenguas: resultados de la investigación en dos comunidades bilingües. *Teoría de la Educación en la Sociedad de la Información*, 12(3), 200-214.
- Ruiz de Zarobe, Y., & Jiménez Catalán, R. M. (2010). *Content and language integrated learning: Evidence from research in Europe*. Bristol: Multilingual Matters.
- Ruiz de Zarobe, Y., & Lasagabaster, D. (2010). Introduction. In Y. Ruiz de Zarobe & D. Lasagabaster (Eds.), *CLIL in Spain: Implementation, results and teacher training* (pp. ix-xvii). Newcastle upon Tyne: Cambridge Scholars Publishing.
- Ruiz de Zarobe, Y., & Lasagabaster, D. (Eds.) (2010). *CLIL in Spain: Implementation, results and teacher training*. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Rumlach, D. (2014). Prospective CLIL and non-CLIL students' interest in English (classes): a quasi-experimental study on German sixth-graders. In R. Breeze, C. Llamas Saíz, C. Martínez Pasamar, & C. Taberero Sala (Eds.), *Integration of theory and practice in CLIL* (pp. 75-95). Amsterdam: Rodopi.

- Rumlich, D. (2016). *Evaluating bilingual education in Germany: CLIL students' general English proficiency, EFL self-concept and interest*. Frankfurt am Main: Peter Lang.
- Rumlich, D. (2017). CLIL theory and empirical reality –Two sides of the same coin? A quantitative-longitudinal evaluation of general EFL competence and affective-motivational dispositions in CLIL students at German secondary schools. *Journal of Immersion and Content-Based Language Education*, 5, 110–134. <https://doi.org/10.1075/jic.5.1.05rum>
- Ryan, S. (2008). *The Ideal L2 Selves of Japanese Learners of English* (Unpublished doctoral dissertation). University of Nottingham, Nottingham, England.
- Ryan, S. (2009). Self and identity in L2 motivation in Japan: The ideal L2 self and Japanese learners of English. In Z. Dörnyei, & E. Ushioda (Eds.), *Motivation, Language Identity and the L2 Self* (pp. 120-143). Bristol: Multilingual Matters.
- Ryan, S., & Mercer, S. (2011). Natural talent, natural acquisition and abroad: Learner attributions of agency in language learning. In G. Murray, X. Gao & T. Lamb (Eds.), *Identity, Motivation and Autonomy in Language Learning* (pp. 141-154). Bristol: Multilingual Matters.
- Salcedo, P., Del Valle, M., Quintanilla, A., & Zambrano, C. (2016, November). Lexical Availability of Students in Mathematics Education in Southern Chile. Paper presented at *9th Annual International Conference of Education, Research and Innovation* (pp. 3676-3683). Seville, Spain.
- Salcedo, P., Ferreira, A., del Valle, M., Cerda, G., & Friz, M. (2014). *Plataforma adaptativa online para el fortalecimiento de las competencias matemáticas y pedagógicas a partir del estudio léxico semántico de estudiantes y profesores de Pedagogía en Matemática*. Proyecto Fondecyt N° 1140457, Comisión Nacional de Investigación Científica y Tecnológica, Gobierno de Chile.
- Samper Hernández, M. (2014). The relationship of language proficiency to the lexical availability of learners of Spanish. In R. M. Jiménez Catalán (Ed.), *Lexical Availability in English and Spanish as a Second Language* (pp. 103-124). Dordrecht: Springer.

- Samper Hernández, M., & Jiménez Catalán, R. M. (2014). Researching Lexical Availability in L2: Some Methodological Issues. In R. M. Jiménez Catalán (Ed.), *Lexical Availability in English and Spanish as a Second Language* (pp. 189-205). Dordrecht: Springer.
- San Isidro, X. (2010). An Insight into Galician CLIL. In D. Lasagabaster & Y. Ruiz de Zarobe (Eds), *CLIL in Spain: Implementation, Results and Teacher Training* (pp. 55-78). Newcastle upon Tyne: Cambridge Scholars Publishing.
- San Isidro, X. (2018). Innovations and challenges in CLIL implementation in Europe. *Theory into Practice*, 57(3), 185-195.
- San Isidro, X. (2019). The multi-faceted effects of CLIL: a literature review. *NEXUS (AEDEAN)*, 1, 33-49.
- San Isidro, X., & Lasagabaster, D. (2018). The impact of CLIL on pluriliteracy development and content learning in a rural multilingual setting: A longitudinal study. *Language Teaching Research*, 23(5), 584-602.
- Sandu, B. M., & Oxbrow, G. L. (2021). Exploring the L2 motivational self system through lexical availability and overseas experience amongst English-major students in Gran Canaria. *RAEL. Revista Electrónica de Lingüística Aplicada*. 19(2), 108-125.
- Sangeetha, T., & Mekala, S. (2021). English Language Assessment and Grading System in Higher Education: An Overview. *IUP Journal of Soft Skills*, 15(4), 17-28.
- Santos Díaz, I. C. (2015). La influencia del conocimiento de varias lenguas extranjeras en el léxico disponible. In A. Gordejuela Senosiáin, D. Izquierdo Alegría, F. Jiménez Berrio, A. de Lucas Vicente & M. Casado Velarde (Eds.), *Lenguas, lenguaje y lingüística. Contribuciones desde la Lingüística General* (pp. 477-484). Pamplona: Servicio de Publicaciones de la Universidad de Navarra.
- Santos Díaz, I. C. (2017a). Incidencia de la lectura en el vocabulario en lengua materna y extranjera. *Ocnos: Revista de Estudios Sobre Lectura*, 16(1), 79-88.
- Santos Díaz, I. C. (2017b). Organización de las palabras en la mente en lengua materna y lengua extranjera (inglés y francés). *Pragmalingüística*, 25, 603-617

- Santos Díaz, I. C. (2017c). Selección del léxico disponible: propuesta metodológica con fines didácticos. *Porta Linguarum: Revista Internacional de Didáctica de las Lenguas Extranjeras*, 27, 122-139.
- Sasaki, M. (2011). Effects of varying lengths of study-abroad experiences on Japanese EFL students' L2 writing ability and motivation: A Longitudinal Study. *TESOL Quarterly*, 45(1), 81-105.
- Sasaki, M., Kozaki, Y., & Ross, S. J. (2017). The impact of normative environments on learner motivation and L2 reading ability growth. *The Modern Language Journal*, 101(1), 163–178.
- Saville-Troike, M., & Barto, K. (2017). *Introducing Second Language Acquisition* (3rd ed.). Cambridge, UK: Cambridge University Press.
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 17–45.
- Schmidt, R., Boraie, D., & Kassabgy, O. (1996). Foreign language motivation: Internal structure and external connections. *University of Hawai'i Working Papers in English as a Second Language*, 14(2), 1-72.
- Schmitt, N. (1995). A fresh approach to vocabulary: using a word knowledge framework. *RELC Journal*, 26(1), 86-94.
- Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge, UK: Cambridge University Press.
- Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. New York: Palgrave Macmillan.
- Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows. *Language Learning*, 64(4), 913–951.
- Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, 18(1), 55-88.
- Schumann, J. H. (1978). The acculturation model for second language acquisition. In R. C. Gingras (Ed.), *Second language acquisition and foreign language teaching* (pp. 27-50). Arlington, VA: Center for Applied Linguistics.

- Schumann, J. H. (1986). Research on the acculturation model for second language acquisition. *Journal Of Multilingual & Multicultural Development*, 7(5), 379-392.
- Seashore, R. H., & Eckerson, L. D. (1940). The measurement of individual differences in general English vocabularies. *Journal of Educational Psychology*, 31(1), 14.
- Sędek, G. (1995). *Bezradność intelektualna w szkole*. Wydawnictwo Instytutu Psychologii.
- Segalowitz, N., & Freed, B. F. (2004). Context, contact, and cognition in oral fluency acquisition: Learning Spanish in at home and study abroad contexts. *Studies in Second Language Acquisition*, 26(2), 173-199.
- Segalowitz, N., Freed, B., Collentine, J., Lafford, B., Lazar, N., & Díaz-Campos, M. (2004). A comparison of Spanish second language acquisition in two different learning contexts: Study abroad and the domestic classroom. *Frontiers: The Interdisciplinary Journal of Study Abroad*, 10, 1-18.
- Seikkula-Leino, J. (2007). CLIL learning: Achievement levels and affective factors. *Language and Education*, 21(4), 328-341.
- Serwatka, M., & Healey, A. E. (1998). On the status of the count-mass distinction in a mental grammar. In A. E. Healy & L. E. Bourne, Jr. (Eds.), *Foreign language learning: Psycholinguistic studies on training and retention* (pp. 113-140). Mahwah, NJ: Lawrence Erlbaum.
- Sierra, J. (1994). Modelos de enseñanza bilingüe: sus resultados y su futuro. In M. Siguán (Ed.), *Las lenguas en la escuela* (pp. 61-76). Barcelona: ICE from Universitat de Barcelona.
- Šifrar Kalan, M. (2014). Slovene students' lexical availability in English and Spanish. In R. M. Jiménez Catalán (Ed.), *Lexical Availability in English and Spanish as a Second Language* (pp. 125-138). Dordrecht: Springer.
- Šifrar Kalan, M. (2015). Lexical Availability and L2 Vocabulary Acquisition. *Journal of Foreign Language Teaching and Applied Linguistics*, 2(2), 191-200.
- Singleton, D., & Singleton, E. (2002). The C-Test and L2 acquisition/processing research. In J. A. Coleman, R. Grotjahn & U. Raatz (Eds.), *University Language Testing and The C-Test* (pp. 143-168). Bochum: AKS-Verlag.

- Skehan, P. (1996). Second language acquisition research and task-based instruction. In J. Willis & D. Willis (Eds.), *Challenge and change in language teaching* (pp. 17-30). Oxford: Heinemann.
- Somers, T., & Llinares, A. (2018). Students' motivation for content and language integrated learning and the role of programme intensity. *International Journal of Bilingual Education and Bilingualism*, 23(10), 1-16.
- Somers, T., & Surmont, J. (2012). CLIL and immersion: how clear-cut are they? *ELT Journal*, 66(1), 113-116.
- Spiri, J. (2008). Online study of frequency list vocabulary with the WordChamp website. *Reflections on English Language Teaching*, 7(1), 21-36.
- Stemmer, B. (1991). *What's on a C-Test taker's mind? Mental processes in C-Test taking*. Bochum: Brockmeyer.
- Stubbe, R. (2015). Replacing translation tests with Yes/No tests. *Vocabulary Learning and Instruction*, 4(2), 38-48.
- Sugita McEown, M., & Takeuchi, O. (2014). Motivational strategies in EFL classrooms: how do teachers impact students' motivation? *Innovation in Language Learning and Teaching*, 8(1), 20-38.
- Sugita McEown, M., Sawaki, Y., & Harada, T. (2017). Foreign language learning motivation in the Japanese context: Social and political influences on self. *The Modern Language Journal*, 101(3), 533-547.
- Sunderland, J. (2010). Theorizing gender perspectives in foreign and second language learning. In R. M. Jiménez Catalán (Ed.), *Gender Perspectives on Vocabulary in Foreign and Second languages* (pp. 1-19). New York: Palgrave Macmillan.
- Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. M. Gass & C. G. Madden (Eds.), *Input in Second Language Acquisition* (pp. 235-253). Rowley, MA: Newbury House.
- Sweet, H. (1899). *The Practical Study of Languages*. London: Dent. (Republished 1964, London: Oxford University Press).

- Sylvén, L. K. (2004). *Teaching in English or English Teaching? On the Effects of Content and Language Integrated Learning on Swedish Learners' Incidental Vocabulary Acquisition*. (Unpublished doctoral dissertation). Goteborg University.
- Sylvén, L. K. (2006). How is Extramural Exposure to English among Swedish School Students Used in the CLIL Classroom? *VIEWZ Vienna English Working Papers (Special Issue: Current Research on CLIL)*. 15(3), 47–54.
- Sylvén, L. K. (2019). Extramural English. In L. K. Sylvén (Ed.), *Investigating Content and Language Integrated Learning: Insights from Swedish High Schools* (pp.152-167). Bristol: Multilingual Matters.
- Sylvén, L. K., & Ohlander, S. (2014). The CLISS project: Receptive vocabulary in CLIL versus non-CLIL groups. *Moderna Språk*, 108(2), 80-114.
- Sylvén, L. K., & Thompson, A. S. (2015). Language learning motivation and CLIL: Is there a connection? *Journal of Immersion and Content-Based Language Education*, 3(1), 28-50.
- Tabuenca Cuevas, M. F., & Alcaraz Mármol, G. (2014). From EFL to CLIL teaching and learning vocabulary in the primary classroom. In J. R. Calvo-Ferrer & M. A. Campos-Pardillos (Eds.), *Investigating lexis vocabulary teaching, ESP, lexicography and lexical innovation* (pp. 27-37). Newcastle Upon Tyne: Cambridge Scholars Publishing.
- Taguchi, T., Magid, M., & Papi, M. (2009). The L2 motivational self system among Japanese, Chinese and Iranian learners of English: A comparative study. In Z. Dörnyei, & E. Ushioda (Eds.), *Motivation, Language Identity and the L2 Self* (pp. 66-97). Bristol: Multilingual Matters.
- Talbot, M. (2019). *Language and Gender* (3rd ed.). Cambridge: Polity Press.
- Tanaka, M. (2017). Examining EFL vocabulary learning motivation in a demotivating learning environment. *System*, 65, 130-138.
- Tankó, G. (2021). Gender-related differences in performing a test-task in academic writing: Insights from performance data on an argumentative essay task. In É. Illés & J. Kenyeres (Eds.), *Changing Perspectives: Studies in English* (pp. 301-326). Budapest: Eötvös Loránd University.

- Tankó, G., & Csizér, K. (2014). Investigating English majors' individual differences through their argumentative essays. In V. Ruttkay & B. Gárdos (Eds.), *Proceedings of the 11th Conference of the Hungarian Society for the Study of English* (pp. 701–718). Budapest: L'Harmattan Publishing House.
- Templin M. C. (1957). *Certain language skills in children*. Minneapolis: University of Minnesota Press.
- Thompson, A. S., & Erdil-Moody, Z. (2016). Operationalizing multilingualism: language learning motivation in Turkey. *International Journal of Bilingual Education and Bilingualism*, 19(3), 314-331.
- Thompson, G., & McKinley, J. (2018). Integration of content and language learning. In J. L. Lontas M. DelliCarpini, & S. Abrar-ul-Hassan (Eds.), *The TESOL Encyclopedia of English Language Teaching*. John Wiley & Sons. <https://doi.org/10.1002/9781118784235>
- Ting, Y. L. T. (2010). CLIL appeals to how the brain likes its information: examples from CLIL-(Neuro) Science. *International CLIL Research Journal*, 1(3), 3-18.
- Torruella, J., & Capsada, R. (2013). Lexical statistics and tipological structures: a measure of lexical richness. *Procedia-Social and Behavioral Sciences*, 95, 447-454.
- Tragant, E., Marsol, A., Serrano, R., & Llanes, À. (2016). Vocabulary learning at primary school: a comparison of EFL and CLIL. *International Journal of Bilingual Education and Bilingualism*, 19(5), 579-591.
- Treffers-Daller, J. (2013). Measuring lexical diversity among L2 learners of French: an exploration of the validity of D, MTLD and HD-D as measures of language ability. In S. Jarvis & M. Daller (Eds.), *Vocabulary knowledge: human ratings and automated measures* (pp. 79-104). Amsterdam: Benjamins.
- Treffers-Daller, J., Parslow, P., & Williams, S. (2018). Back to basics: How measures of lexical diversity can help discriminate between CEFR levels. *Applied Linguistics*, 39(3), 302-327.
- Trenchs-Parera, M., & Juan-Garau, M. (2014). A longitudinal study of learners' motivation and beliefs in at home and study abroad contexts. In C. Pérez-Vidal

- (Ed.), *Language Acquisition in Study Abroad and Formal Instruction Contexts* (pp. 259-282). Amsterdam: Benjamins.
- Tsuchiya, K., & Pérez Murillo, D. (Eds.) (2019). *Content and Language Integrated Learning in Spanish and Japanese Contexts*. Cham: Palgrave Macmillan.
- Tuero, C., González-Boto, R., Espartero, J., & Zapico, B. (2014). Gender stereotypes as a determinant of participation in sports in childhood. *Science and Sports*, 29, S20.
- Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, 92(1-2), 231-270.
- Ure, J. (1971). Lexical density and register differentiation. In G. Perren & J. Trim (Eds.), *Applications of Linguistics* (pp. 443-452). London: Cambridge University Press:
- Ushioda, E. (1994). L2 motivation as a qualitative construct. *Teanga*, 14, 76– 48.
- Ushioda, E. (1996). Developing a dynamic concept of motivation. In T. J. Hickey (Ed.), *Language, education and society in a changing world* (pp. 239–245). Clevedon, England: Multilingual Matters.
- Ushioda, E. (1998). Effective motivational thinking: A cognitive theoretical approach to the study of language learning motivation. In E. A. Soler & V. C. Espurz (Eds.), *Current issues in English language methodology* (pp. 77–89). Castelló de la Plana, Spain: Universitat Jaume.
- Ushioda, E. (2001). Language learning at university: Exploring the role of motivational thinking. In Z. Dörnyei & R. Schmidt (Eds.), *Motivation and second language acquisition* (pp. 91–124). Honolulu: University of Hawaii Press.
- Ushioda, E. (2009). A person-in-context relational view of emergent motivation, self and identity. In Z. Dörnyei & E. Ushioda (Eds), *Motivation, Language Identity and the L2 Self* (pp. 215-228). Bristol: Multilingual Matters.
- Ushioda, E. (2014). Context and complex dynamic systems theory. In Z. Dörnyei, P. D. MacIntyre & A. Henry (Eds.), *Motivational Dynamics in Language Learning* (pp. 85–96). Bristol: Multilingual Matters.

- Ushioda, E., & Dörnyei, Z. (2012). Motivation. In S. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 396-409). New York: Routledge.
- Ushioda, E., & Dörnyei, Z. (2017). Beyond global English: Motivation to learn languages in a multicultural world: Introduction to the special issue. *The Modern Language Journal*, 101(3), 451-454.
- van der Slik, F. W., van Hout, R. W., & Schepens, J. J. (2015). The gender gap in second language acquisition: Gender differences in the acquisition of Dutch among immigrants from 88 countries with 49 mother tongues. *PloS One*, 10(11).
- van Geert, P. (2011). The contribution of complex dynamic systems to development. *Child Development Perspectives*, 5(4), 273-278.
- van Ginkel, C. I., & van der Linden, E. H. (1996). Word associations in foreign language learning and foreign language loss. In K. Sajavaara & C. Fairweather (Eds.), *Approaches to second language acquisition* (pp. 25-35). Jyväskylä: University of Jyväskylä.
- van Mensel, L., Hiligsmann, P., Mettwie, L., & Galand, B. (2020). CLIL, an elitist language learning approach? A background analysis of English and Dutch CLIL pupils in French-speaking Belgium. *Language, Culture and Curriculum*, 33(1), 1-14.
- Vermeer, A. (2001). Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. *Applied Psycholinguistics*, 22(2), 217-234.
- Vermeer, A. (2004). The relation between lexical richness and vocabulary size in Dutch L1 and L2 children. In P. Boogards & B. Laufer (Eds.), *Vocabulary in a second language: Selection, Acquisition, and Testing* (pp. 173-189). Amsterdam: Benjamins.
- Verspoor, M., de Bot, K., & Xu, X. (2015). The effects of English bilingual education in the Netherlands. *Journal of Immersion and Content-Based Language Education*, 3(1), 4-27.
- Vež, J. M. (2011). Plurilingual education in bilingual areas: The case of Galicia. In D. Madrid & S. Hughes (Eds.), *Studies in Bilingual Education* (pp. 327- 350). Frankfurt am Main: Peter Lang.

- Villabona, N., & Cenoz, J. (2021). The integration of content and language in CLIL: a challenge for content-driven and language-driven teachers. *Language, Culture and Curriculum*, 35(1), 1-15.
- Walstad, W. B., & Robson, D. (1997). Differential item functioning and male-female differences on multiple-choice tests in economics. *The Journal of Economic Education*, 28(2), 155-171.
- Walters, J. (2012). Aspects of validity of a test of productive vocabulary: Lex30. *Language Assessment Quarterly*, 9(2), 172-185.
- Wang, T., & Liu, Y. (2020). Dynamic L3 selves: a longitudinal study of five university L3 learners' motivational trajectories in China. *The Language Learning Journal*, 48(2), 201-212.
- Waring, R. (1997). A comparison of the receptive and productive vocabulary sizes of some second language learners. *Immaculata*, 1(1), 53-68.
- Webb, S. (2008). Receptive and productive vocabulary sizes of L2 learners. *Studies in Second language acquisition*, 30(1), 79-95.
- Webb, S. A., & Sasao, Y. (2013). New directions in vocabulary testing. *RELC Journal*, 44(3), 263-277.
- Wechsler, D. (1949). *Wechsler Intelligence Scale for Children*. New York: The Psychological Corporation.
- Wei, L., & Moyer, M. G. (2009). *The Blackwell Guide to Research Methods in Bilingualism and Multilingualism*. Oxford: Blackwell.
- Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review*, 92(4), 458-573.
- Weiner, B. (2013). *Human motivation*. New Jersey: Lawrence Erlbaum Associates.
- Wesche, M., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: Depth versus breadth. *Canadian Modern Language Review*, 53(1), 13-40.
- West, M. P. (1953). *A General Service List of English Words*. London: Longman, Green & Co.

- Wiesemes, R. (2009). Developing theories of practices in CLIL: CLIL as post-method pedagogies. In R. M. Jiménez Catalán & Y. Ruiz de Zarobe (Eds.), *Content and language integrated learning: Evidence from research in Europe* (pp. 41-59). Bristol: Multilingual Matters.
- Wilkins, D. A. (1972). *Linguistics in language teaching*. London: Arnold.
- Williams, M., & Burden, R. L. (1997). *Psychology for Language Teachers*. Cambridge: Cambridge University Press.
- Woodrow, L. (2017). Motivation in language learning. In R. Breeze & C. S. Guinda (Eds.), *Essential competencies for English-medium university teaching* (pp. 235-248). Cham: Springer.
- Word-association test (n.d.). In *Collins Dictionary*. Retrieved from <https://www.collinsdictionary.com/es/diccionario/ingles/word-association-test>
- Xanthou, M. (2010). Current Trends in L2 Vocabulary Learning and Instruction. Is CLIL the right approach? In A. Psaltou-Joycey & M Mattheoudakis (Eds.), *14th International Conference: Advances in Research on Language Acquisition and Teaching – Selected Papers* (pp. 459-471). Thessaloniki, Greece: Greek Applied Linguistics Association (GALA).
- Yashima, T. (2009). International posture and the ideal L2 self in the Japanese EFL context. In Z. Dörnyei, & E. Ushioda (Eds.), *Motivation, Language Identity and the L2 Self* (pp. 144-163). Bristol: Multilingual Matters.
- Yashima, T. (2019). L2 Motivation and Willingness to Communicate. In M. Lamb, K. Csizér, A. Henry & S. Ryan (Eds.), *The Palgrave Handbook of Motivation for Language Learning* (pp. 203-223). Cham: Palgrave Macmillan.
- Yim, O., Clément, R., & MacIntyre, P. D. (2019). The contexts of SLA motivation: linking ideologies to situational variations. In M. Lamb, K. Csizér, A. Henry & S. Ryan (Eds.), *The Palgrave Handbook of Motivation for Language Learning* (pp. 225-244). Cham: Palgrave Macmillan.
- Young, R. (2018). What is the effect of Content Language Integrated Learning on pupil motivation to learn a foreign language? *Scottish Languages Review & Digest*, 34, 29-38.

- Yue, Y., & Fan, S. (2016). Measurement of Vocabulary Knowledge: Problems and Solutions. In S. Fan & J. Fielding-Wells (Eds.), *What is Next in Educational Research?* (pp. 171-182). Rotterdam, The Netherlands: Sense Publishers.
- Zajonc, R. B. (1968). Attitudinal effects of mere exposures. *Journal of Personality and Social Psychology*, 9(2, Pt. 2), 1-27.
- Zajonc, R. B. (2001). Mere exposure: A gateway to the subliminal. *Current directions in psychological science*, 10(6), 224-228.
- Zaker, A. (2016). The acculturation model of second language acquisition: Inspecting weaknesses and strengths. *Indonesian EFL Journal*, 2(2), 80-87.
- Zhang, Q. (2007). Teacher misbehaviors as learning demotivators in college classrooms: A cross-cultural investigation in China, Germany, Japan, and the United States. *Communication Education*, 56(2), 209-227.
- Zhang, X. (2019). Foreign language anxiety and foreign language performance: A meta-analysis. *The Modern Language Journal*, 103(4), 763-781.
- Zhang, X. (2020). "Don't You Reckon That Being a Multilingual Person Is Cool?"—Investigating the L3 Learning Motivation of Chinese University Learners. *Open Journal of Social Sciences*, 8(08), 453-461.
- Zhang, X., & Lu, X. (2014). A longitudinal study of receptive vocabulary breadth knowledge growth and vocabulary fluency development. *Applied Linguistics*, 35(3), 283-304.
- Zheng, Y. (2016). The complex, dynamic development of L2 lexical use: A longitudinal study on Chinese learners of English. *System*, 56, 40-53.
- Zydati, W. (2007). *Deutsch-Englische Zge in Berlin (DEZIBEL): Eine Evaluation des bilingualen Sachfachunterrichts an Gymnasien; Kontext, Kompetenzen, Konsequenzen* [German-English education in Berlin. An evaluation of bilingual education in upper secondary school: Context, competences, consequences]. Frankfurt am Main: Peter Lang.
- Zydati, W. (2012). Linguistic thresholds in the CLIL classroom? The threshold hypothesis revisited. *International CLIL Research Journal*, 1(4), 17-28.

Appendices

Appendix A: English C-test

Participante: _____

Esta parte consta de **4 textos con palabras** a las que les **faltan algunas letras**. La finalidad de la prueba es completar el máximo número de palabras. Para completarlas hay que añadir un número **igual** de letras, **o una más o una menos** (si es una palabra con un número **impar** de letras) a las que ya vienen dadas.

Ejemplo:

School for heroes

The internationally acclaimed Fire Services College in Gloucestershire is widely regarded as the best fire college in the world. The site, a former airfield which once launched Wellington bombers for raids on Germany, was recently transferred from the Home Office to the Fire Services. And such is its reputation that people come from as far away as China and Trinidad to sharpen up their leadership skills. Even fire chiefs from the other side of the world have attended as College students.

1. Geography

The UK is located on a group of islands known as the British Isles, which lie between the Atlantic Ocean and the North Sea, northwest of France. At i _____ widest t _____ UK i _____ 300 mi _____ across a _____ 600 mi _____ from No _____ to So _____. It sha _____ a sin _____ land bor _____ with the Irish Repu _____. Despite i _____ relatively sm _____ size t _____ UK boa _____ incredibly var _____ and of _____ very beau _____ scenery, fr _____ the mountains and valleys of the North and West to the rolling landscape of the South, and from downland and heath to fens and marshland.

2. UK Passport Service

A new passport office that has opened in London will help the UK Passport Service provide a much better service to customers who need a passport urgently. The new office runs on an appointment-only basis, removing the need for a lengthy wait before being seen. The new building, Globe House replaced the Petty France office, which after 50 years of continuous service, has now closed its doors. The London Passport Office has the capacity to issue 5000 passports weekly.

3. Record employment

Latest employment figures show that there are 28.2 million people in work. Work & Pensions Secretary Alistair Darling said this showed the UK labour market has coped well so far with the current international economic uncertainty. Mr Darling said: "Employment continues to rise, with this month's figures showing a record 28.2 million people in work. There are 65,000 more people in work than last quarter and 252,000 more than last year. Although both measures of unemployment have risen slightly, they are still significantly lower than they were a year ago." The latest claimant count figures, for the month on Dec 13 2001, show a rise of 3,200 on the previous month. At 963,500 claimants, it remains 70,000 lower than this time last year.

4. Government consults on plans to modernise animal welfare

Plans to review, modernise and simplify outdated laws on animal welfare have been announced by the Government. Animal welfare groups, local authority representatives, courts, police and industry are to be consulted in which will be a further reaching review drawing together the environmental and industry concerns over animal welfare. The Department for the Environment, Food and Rural Affairs (DEFRA) wants to hear views on the existing 11 Acts of Parliament governing the welfare of pets and farm animals.

Appendix B: French C-test

Participante: _____

Esta parte consta de **4 textos** con **palabras** a las que les **faltan algunas letras**. La finalidad de la prueba es completar el máximo número de palabras. Para completarlas hay que añadir un número **igual** de letras, **o una más o una menos** (si es una palabra con un número **impar** de letras) a las que ya vienen dadas.

Ejemplo:

Difficile mesure du recul des glaces

En climatologie, les bonnes nouvelles peuvent ne pas s'avérer si bonnes qu'elles en ont l'air. Ainsi de celles apportées par les travaux franco-canadiens publiés dimanche 17 janvier dans la revue Nature Geoscience, qui revoient à la baisse les estimations précédentes de la fonte moyenne des glaciers de l'Alaska depuis un demi-siècle.

1. Strauss-Kahn exhorte les Etats à agir vite

Les réunions financières de Washington vendredi vont permettre de faire le point sur la crise et de tester la stratégie adoptée lors du récent G20 de Londres. Si l'o_____ a enc_____ un do_____ sur l_____ montée e_____ puissance d_____ pays émerg_____ dans l_____ gouvernance écono_____ mondiale, i_____ sera définit_____ balayé cet_____ semaine. U_____ G20 d_____ ministres d_____ Finances e_____ des gouver_____ de banq_____ centrales d_____ vingt pa_____ les pl_____ importants économiquement va se tenir vendredi à Washington, à l'initiative du secrétaire américain au Trésor américain, Timothy Geithner. La réunion donnera le coup d'envoi aux assemblées de printemps du Fonds monétaire international et de la Banque mondiale (25 et 26 avril).

2. Les vaches folles

Selon une récente enquête, 45 % des Français auraient diminué ou cessé de manger de la viande de boeuf depuis le début de la crise de la vache folle. Ils se tournent vers les viandes blanches et la nourriture végétale. On court sans doute davantage de risques en prenant le volant de sa voiture qu'en consommant une entrepaille. Mais, comme le remarque dernièrement un sociologue, «le Français veut bien mourir en conduisant mais pas en mangeant».

3. La grève s'atténue sur le réseau Paris-Nord

Le mouvement de grève lancé hier par les conducteurs et contrôleurs de la SNCF officiant sur les lignes K et H du réseau Paris-Nord devrait s'affaiblir aujourd'hui. Hier matin, un train sur deux roulait sur la ligne K, contre un train sur trois sur la ligne H. Les grévistes ont reconstruit le mouvement au cours d'une assemblée générale, bien que celui-ci ne semble pas avoir été fortement suivi. La grève, décidée pour des motifs salariaux et pour des questions de notation, devrait encore s'affaiblir aujourd'hui, selon la SNCF, qui prévoit une reprise normale du trafic sur la portion K.

4. L'importance de se faire vacciner

La semaine européenne de la vaccination vient de commencer hier. Une semaine de plus consacrée à un problème de santé par lequel les autorités de santé tirent aujourd'hui la sonnette d'alarme. Les vaccins ne sont pas assez faits dans notre pays et certaines maladies que l'on pensait disparues réapparaissent comme la rougeole par exemple... On est passé à près de 600 cas en 2008, ce qui, compte tenu du non-respect de l'obligation de déclarer cette maladie, signifie que plusieurs milliers de cas soient survenus.

Appendix C: English Lexical Availability Task

Participante: _____

PRUEBA DE DISPONIBILIDAD LÉXICA INGLÉS

.....

INSTRUCCIONES:

- En esta prueba hay 5 enunciados.
- Escribe **en 2 minutos** las palabras que te sugiera cada enunciado siguiendo el orden numérico.
- Escribe las palabras **en inglés**.
- El profesor/a te irá marcando el tiempo para cada palabra.
- Una vez acabado el tiempo no puedes incluir más palabras.
- Por favor, escribe con letra legible.

ANIMALS

1. _____
2. _____
3. _____
4. _____
5. _____
6. _____
7. _____
8. _____
9. _____
10. _____
11. _____
12. _____
13. _____
14. _____
15. _____
16. _____
17. _____
18. _____
19. _____
20. _____

21. _____
22. _____
23. _____
24. _____
25. _____
26. _____
27. _____
28. _____
29. _____
30. _____
31. _____
32. _____
33. _____
34. _____
35. _____
36. _____
37. _____
38. _____
39. _____
40. _____

FOOD AND DRINK

1. _____
2. _____
3. _____
4. _____
5. _____
6. _____
7. _____
8. _____
9. _____
10. _____
11. _____
12. _____
13. _____
14. _____
15. _____
16. _____
17. _____
18. _____
19. _____
20. _____

21. _____
22. _____
23. _____
24. _____
25. _____
26. _____
27. _____
28. _____
29. _____
30. _____
31. _____
32. _____
33. _____
34. _____
35. _____
36. _____
37. _____
38. _____
39. _____
40. _____

SPORTS AND PHYSICAL ACTIVITIES

1. _____
2. _____
3. _____
4. _____
5. _____
6. _____
7. _____
8. _____
9. _____
10. _____
11. _____
12. _____
13. _____
14. _____
15. _____
16. _____
17. _____
18. _____
19. _____
20. _____

21. _____
22. _____
23. _____
24. _____
25. _____
26. _____
27. _____
28. _____
29. _____
30. _____
31. _____
32. _____
33. _____
34. _____
35. _____
36. _____
37. _____
38. _____
39. _____
40. _____

ENVIRONMENT AND CLIMATE

1. _____
2. _____
3. _____
4. _____
5. _____
6. _____
7. _____
8. _____
9. _____
10. _____
11. _____
12. _____
13. _____
14. _____
15. _____
16. _____
17. _____
18. _____
19. _____
20. _____

21. _____
22. _____
23. _____
24. _____
25. _____
26. _____
27. _____
28. _____
29. _____
30. _____
31. _____
32. _____
33. _____
34. _____
35. _____
36. _____
37. _____
38. _____
39. _____
40. _____

ECONOMY AND MONEY

1. _____
2. _____
3. _____
4. _____
5. _____
6. _____
7. _____
8. _____
9. _____
10. _____
11. _____
12. _____
13. _____
14. _____
15. _____
16. _____
17. _____
18. _____
19. _____
20. _____

21. _____
22. _____
23. _____
24. _____
25. _____
26. _____
27. _____
28. _____
29. _____
30. _____
31. _____
32. _____
33. _____
34. _____
35. _____
36. _____
37. _____
38. _____
39. _____
40. _____

Appendix D: French Lexical Availability Task

Participante: _____

PRUEBA DE DISPONIBILIDAD LÉXICA FRANCÉS

.....

INSTRUCCIONES:

- En esta prueba hay 5 enunciados.
- Escribe **en 2 minutos** las palabras que te sugiera cada enunciado siguiendo el orden numérico.
- Escribe las palabras **en francés**.
- El profesor/a te irá marcando el tiempo para cada palabra.
- Una vez acabado el tiempo no puedes incluir más palabras.
- Por favor, escribe con letra legible.

LES ANIMAUX

1. _____
2. _____
3. _____
4. _____
5. _____
6. _____
7. _____
8. _____
9. _____
10. _____
11. _____
12. _____
13. _____
14. _____
15. _____
16. _____
17. _____
18. _____
19. _____
20. _____

21. _____
22. _____
23. _____
24. _____
25. _____
26. _____
27. _____
28. _____
29. _____
30. _____
31. _____
32. _____
33. _____
34. _____
35. _____
36. _____
37. _____
38. _____
39. _____
40. _____

LA NOURRITURE ET LES BOISSONS

1. _____

2. _____

3. _____

4. _____

5. _____

6. _____

7. _____

8. _____

9. _____

10. _____

11. _____

12. _____

13. _____

14. _____

15. _____

16. _____

17. _____

18. _____

19. _____

20. _____

21. _____

22. _____

23. _____

24. _____

25. _____

26. _____

27. _____

28. _____

29. _____

30. _____

31. _____

32. _____

33. _____

34. _____

35. _____

36. _____

37. _____

38. _____

39. _____

40. _____

LE SPORT ET LES ACTIVITÉS PHYSIQUES

- | | |
|-----------|-----------|
| 1. _____ | 21. _____ |
| 2. _____ | 22. _____ |
| 3. _____ | 23. _____ |
| 4. _____ | 24. _____ |
| 5. _____ | 25. _____ |
| 6. _____ | 26. _____ |
| 7. _____ | 27. _____ |
| 8. _____ | 28. _____ |
| 9. _____ | 29. _____ |
| 10. _____ | 30. _____ |
| 11. _____ | 31. _____ |
| 12. _____ | 32. _____ |
| 13. _____ | 33. _____ |
| 14. _____ | 34. _____ |
| 15. _____ | 35. _____ |
| 16. _____ | 36. _____ |
| 17. _____ | 37. _____ |
| 18. _____ | 38. _____ |
| 19. _____ | 39. _____ |
| 20. _____ | 40. _____ |

L'ÉNVIRONNEMENT ET LE CLIMAT

1. _____

2. _____

3. _____

4. _____

5. _____

6. _____

7. _____

8. _____

9. _____

10. _____

11. _____

12. _____

13. _____

14. _____

15. _____

16. _____

17. _____

18. _____

19. _____

20. _____

21. _____

22. _____

23. _____

24. _____

25. _____

26. _____

27. _____

28. _____

29. _____

30. _____

31. _____

32. _____

33. _____

34. _____

35. _____

36. _____

37. _____

38. _____

39. _____

40. _____

L'ÉCONOMIE ET L'ARGENT

- | | |
|-----------|-----------|
| 1. _____ | 21. _____ |
| 2. _____ | 22. _____ |
| 3. _____ | 23. _____ |
| 4. _____ | 24. _____ |
| 5. _____ | 25. _____ |
| 6. _____ | 26. _____ |
| 7. _____ | 27. _____ |
| 8. _____ | 28. _____ |
| 9. _____ | 29. _____ |
| 10. _____ | 30. _____ |
| 11. _____ | 31. _____ |
| 12. _____ | 32. _____ |
| 13. _____ | 33. _____ |
| 14. _____ | 34. _____ |
| 15. _____ | 35. _____ |
| 16. _____ | 36. _____ |
| 17. _____ | 37. _____ |
| 18. _____ | 38. _____ |
| 19. _____ | 39. _____ |
| 20. _____ | 40. _____ |

Appendix E: English Language Learning Motivation Questionnaire

Cuestionario: aprendizaje de inglés

A través del siguiente cuestionario pretendemos aproximarnos a algunos aspectos relacionados con el proceso de aprendizaje de inglés. Las respuestas serán anónimas, serán tratadas con discreción y se emplearán para fines exclusivamente científicos. Por ello, le invitamos a que responda con absoluta sinceridad. Muchísimas gracias por su valiosa colaboración.

Parte 1: Datos Personales

1. Edad: _____
2. Género: hombre mujer
3. Nacionalidad: _____
4. Lengua(s) materna(s): _____
5. Lengua(s) extranjera(s) que conoce:

_____ Nivel alto Nivel medio Nivel bajo

_____ Nivel alto Nivel medio Nivel bajo

_____ Nivel alto Nivel medio Nivel bajo
6. Indique con una (X) si ha residido o ha visitado algún país de habla inglesa. En caso afirmativo, especifica cuánto tiempo (días, meses y años aproximadamente).

 No

 Sí _____ días _____ mes(es) _____ año(s) en _____ (lugar)
 _____ días _____ mes(es) _____ año(s) en _____ (lugar)
 _____ días _____ mes(es) _____ año(s) en _____ (lugar)
7. Indique con un número (1-10) la nota final que ha recibido en las últimas evaluaciones.

Nota final: Lengua inglesa _____ Nota final: Economía _____
Nota final: Física y química _____ Nota final: Educación física _____
8. Indique si realice otras actividades en inglés fuera del instituto, p. ej. clases de inglés en una academia.

Actividad: _____ (_____ horas cada semana)
Actividad: _____ (_____ horas cada semana)
Actividad: _____ (_____ horas cada semana)

Parte 2: Indique si está de acuerdo o no con los siguientes enunciados. Señale con una (X) el número que corresponda a su opinión.

1. Estoy totalmente en desacuerdo
2. Estoy en desacuerdo
3. Estoy indeciso/a
4. Estoy de acuerdo
5. Estoy totalmente de acuerdo

	1	2	3	4	5
Estudiar inglés es una pérdida de tiempo.					
Aprendo mucho inglés en el instituto.					
Estudiar inglés es importante para mí porque no quiero que se me considere una persona inculta.					
Aprender inglés es importante para mí porque lo considero un reto en la vida.					
Yo disfruto en una clase de inglés.					
Aprender inglés es importante para mí porque me será imprescindible para conseguir trabajo.					
Si me esfuerzo podré dominar de inglés.					
Seguramente seguiré estudiando inglés después del instituto.					
Si no aprendo inglés no podré trabajar en lo que quiero.					
Aprender educación física en inglés es muy interesante.					
Dada la situación económica en España necesitaré el inglés para trabajar en el extranjero.					
Creo que los demás se reirían de mi inglés.					
Yo disfruto en una clase de física y química.					
Creo que estoy haciendo todo lo que pueda para aprender inglés.					
Estudiar inglés es aburrido.					
No quiero fracasar con el inglés porque mi futuro profesional depende de ello.					
No me puedo imaginar mi futuro sin inglés.					
No hablo inglés por miedo a cometer errores.					
Me encantan las clases de inglés en el instituto.					
Tengo facilidad para aprender inglés.					
Quiero aprender inglés porque la gente a la que quiero piensa que es importante.					
Saber inglés es importante para que se me considere una persona con buena formación.					
Cuando pienso en mi futuro profesional, me veo utilizando el inglés en el trabajo.					
Aprender inglés es importante para mí porque conseguiré un trabajo mejor pagado.					
Yo disfruto en una clase de educación física.					

Preferiría estudiar otro idioma que el inglés.					
Estoy dispuesto a esforzarme mucho en el aprendizaje de inglés.					
Si un extranjero me pidiese indicaciones en la calle, me pondría nervioso/a.					
Aprender inglés es importante para mí porque pienso seguir estudiando en el extranjero.					
Si no fuese por mis seres queridos no aprendería inglés.					
Sueño con dominar el inglés.					
Yo me puedo ver viviendo en el extranjero y desenvolviéndome con la gente en inglés.					
Me pondría muy nervioso/a si tuviese que hablar inglés con un nativo.					
En clase tengo mucho sentido del ridículo al hablar inglés.					
Todos mis compañeros hablan de la importancia de aprender inglés.					
No estudiar inglés tendrá un impacto negativo en mi vida.					
Me puedo ver en una situación en cual estoy hablando inglés con amigos internacionales.					
Tengo intención firme de pasar una temporada en el extranjero para mejorar mi inglés.					
Necesito inglés para la certificación para poder ser docente.					
Aprender inglés es muy interesante.					
Intento aprovechar todo tipo de situaciones para comunicarme en inglés.					
Me encanta como suena el inglés.					
El inglés es muy difícil para mí.					
Estoy trabajando mucho en aprender inglés.					
Aprender economía en inglés es muy interesante.					
En realidad me siento obligado a aprender inglés, no es mi deseo.					
Me encanta escuchar a la gente hablar inglés.					
Tengo profesores muy buenos de inglés.					
Yo disfruto en una clase de economía.					
Mi familia piensa que debería forzarme más con el inglés.					
Aprender inglés es importante para mí porque con ello puedo trabajar a nivel global.					
Tengo mucha curiosidad por la estructura y vocabulario de inglés.					
Aprender física y química en inglés es muy interesante.					
Mi experiencia en clases de inglés siempre ha sido positiva.					
Mis amigos influyen positivamente en mi afán por el inglés.					

¡Muchas gracias por su colaboración!

Appendix F: French Language Learning Motivation Questionnaire

Cuestionario: aprendizaje de francés

A través del siguiente cuestionario pretendemos aproximarnos a algunos aspectos relacionados con el proceso de aprendizaje de francés. Las respuestas serán anónimas, serán tratadas con discreción y se emplearán para fines exclusivamente científicos. Por ello, le invitamos a que responda con absoluta sinceridad. Muchísimas gracias por su valiosa colaboración.

Parte 1: Datos Personales

1. Edad: _____
2. Género: hombre mujer
3. Nacionalidad: _____
4. Lengua(s) materna(s): _____
5. Lengua(s) extranjera(s) que conoce:

_____	<input type="checkbox"/> Nivel alto	<input type="checkbox"/> Nivel medio	<input type="checkbox"/> Nivel bajo
_____	<input type="checkbox"/> Nivel alto	<input type="checkbox"/> Nivel medio	<input type="checkbox"/> Nivel bajo
_____	<input type="checkbox"/> Nivel alto	<input type="checkbox"/> Nivel medio	<input type="checkbox"/> Nivel bajo
6. Indique con una (X) si ha residido o ha visitado algún país de habla francesa. En caso afirmativo, especifica cuánto tiempo (días, meses y años aproximadamente).

No

Sí _____ días _____ mes(es) _____ año(s) en _____ (lugar)

 _____ días _____ mes(es) _____ año(s) en _____ (lugar)

 _____ días _____ mes(es) _____ año(s) en _____ (lugar)
7. Indique con un número (1-10) la nota final que ha recibido en las últimas evaluaciones.

Nota final: Lengua francesa _____ Nota final: Geografía e historia _____
8. Indique si realice otras actividades en francés fuera del instituto, p. ej. clases de francés en una academia.

Actividad: _____ (_____ horas cada semana)

Actividad: _____ (_____ horas cada semana)

Actividad: _____ (_____ horas cada semana)

Parte 2: Indique si está de acuerdo o no con los siguientes enunciados. Señale con una (X) el número que corresponda a su opinión.

6. Estoy totalmente en desacuerdo
7. Estoy en desacuerdo
8. Estoy indeciso/a
9. Estoy de acuerdo
10. Estoy totalmente de acuerdo

	1	2	3	4	5
Me encantan las clases de francés en el instituto.					
No quiero fracasar con el francés porque mi futuro profesional depende de ello.					
Tengo facilidad para aprender francés.					
Creo que los demás se reirían de mi francés.					
Aprender francés es importante para mí porque pienso seguir estudiando en el extranjero.					
Tengo intención firme de pasar una temporada en el extranjero para mejorar mi francés.					
Quiero aprender francés porque la gente a la que quiero piensa que es importante.					
Aprender francés es muy interesante.					
Yo me puedo ver viviendo en el extranjero y desenvolviéndome con la gente en francés.					
Me encanta como suena el francés.					
Cuando pienso en mi futuro profesional, me veo utilizando el francés en el trabajo.					
Si me esfuerzo podré dominar de francés.					
En realidad me siento obligado a aprender francés, no es mi deseo.					
Estudiar francés es una pérdida de tiempo.					
Aprender francés es importante para mí porque me será imprescindible para conseguir trabajo.					
Aprender francés es importante para mí porque conseguiré un trabajo mejor pagado.					
Si un extranjero me pidiese indicaciones en la calle, me pondría nervioso/a.					
Mi experiencia en clases de francés siempre ha sido positiva.					
Intento aprovechar todo tipo de situaciones para comunicarme en francés.					
Estoy trabajando mucho en aprender francés.					
No me puedo imaginar mi futuro sin francés.					
Me puedo ver en una situación en cual estoy hablando francés con amigos internacionales.					
Aprender geografía e historia en francés es muy interesante.					

Preferiría estudiar otro idioma que el francés.					
Sueño con dominar el francés.					
Mi familia piensa que debería forzarme más con el francés.					
Tengo profesores muy buenos de francés.					
Estoy dispuesto a esforzarme mucho en el aprendizaje de francés.					
Aprendo mucho francés en el instituto.					
Saber francés es importante para que se me considere una persona con buena formación.					
No estudiar francés tendrá un impacto negativo en mi vida.					
El francés es muy difícil para mí.					
Necesito francés para la certificación para poder ser docente.					
Dada la situación económica en España necesitaré el francés para trabajar en el extranjero.					
Si no aprendo francés no podré trabajar en lo que quiero.					
Yo disfruto en una clase de geografía e historia.					
Estudiar francés es aburrido.					
Estudiar francés es importante para mí porque no quiero que se me considere una persona inculta.					
No hablo francés por miedo a cometer errores.					
Tengo mucha curiosidad por la estructura y vocabulario de francés.					
Me pondría muy nervioso/a si tuviese que hablar francés con un nativo.					
Mis amigos influyen positivamente en mi afán por el francés.					
Aprender francés es importante para mí porque con ello puedo trabajar a nivel global.					
Aprender francés es importante para mí porque lo considero un reto en la vida.					
En clase tengo mucho sentido del ridículo al hablar francés.					
Yo disfruto en una clase de francés.					
Creo que estoy haciendo todo lo que pueda para aprender francés.					
Todos mis compañeros hablan de la importancia de aprender francés.					
Me encanta escuchar a la gente hablar francés.					
Si no fuese por mis seres queridos no aprendería francés.					
Seguramente seguiré estudiando francés después del instituto.					

¡Muchas gracias por su colaboración!

Leah Geoghegan

Appendix G: School Consent Form

D^a. Leah Geoghegan, doctoranda en la Universidad de la Rioja,

cuya directora de tesis es D^a. María Pilar Agustín Llach,

investigadora del proyecto I+D PGC2018-095260-B-100

“INCREMENTO DEL VOCABULARIO PRODUCTIVO EN INGLES COMO LENGUA
EXTRANJERA: NIVEL, CREATIVIDAD, INPUT”

HACE CONSTAR

Que ha informado a los/as profesores/as de inglés y francés de 4º de la ESO y a la dirección, así como a los alumno/as de este curso del centro _____ sobre los objetivos del proyecto. Ha solicitado su colaboración como centro y ha obtenido permiso para pasar pruebas de inglés y francés como parte del proyecto de tesis doctoral de Leah Geoghegan.

Las pruebas correspondientes se realizarán, de común acuerdo entre profesores de inglés y francés, dirección y equipo de investigación con fecha 14 de febrero de 2020. Al inicio de dicha sesión todos los estudiantes serán informados de la finalidad de nuestra investigación y se les solicitará su colaboración voluntaria y asentimiento.

D^a. Leah Geoghegan se compromete a tratar los datos obtenidos en el centro con total confidencialidad, a informar al centro de los resultados generales de la investigación, así como a utilizar los datos obtenidos en el centro para uso exclusivo de los objetivos propuestos en el proyecto.

Logroño, a 14 de febrero de 2020.

Fdo.:

Fdo.:

Investigador miembro del proyecto I+D

Director/a del centro

PGC2018-095260-B-100

Appendix H: Student Consent Form

Consentimiento Informado para Participantes de Investigación

El propósito de esta ficha de consentimiento es proveer a los participantes en esta investigación con una clara explicación de la naturaleza de la misma, así como de su rol en ella como participantes.

La presente investigación, que se realiza como parte de un proyecto de tesis doctoral, es conducida por Leah Geoghegan, de la Universidad de la Rioja. La meta de este estudio es investigar los efectos del idioma de instrucción en el desarrollo de vocabulario productivo.

Si usted accede a participar en este estudio, se le pedirá responder preguntas en dos pruebas y completar un cuestionario, una vez en inglés y otra vez en francés, lo que supondría unos 50 minutos para cada lengua.

La participación en este estudio es estrictamente voluntaria. La información que se recoja será confidencial y no se usará para ningún otro propósito fuera de los de esta investigación. Sus respuestas a las pruebas y al cuestionario serán codificadas usando un número de identificación y por lo tanto, serán anónimas.

Si tiene alguna duda sobre este proyecto, puede hacer preguntas en cualquier momento durante su participación en él. Igualmente, puede retirarse del proyecto en cualquier momento sin que eso lo perjudique en ninguna forma.

Desde ya le agradecemos su participación.

Acepto participar voluntariamente en esta investigación, conducida por Leah Geoghegan. He sido informado/a de que la meta de este estudio es investigar los efectos del idioma de instrucción en el desarrollo de vocabulario productivo.

Me han indicado también que tendré que responder preguntas en dos pruebas y completar un cuestionario, lo cual tomará aproximadamente 50 minutos para el francés y 50 minutos para el inglés.

Reconozco que la información que yo provea en el curso de esta investigación es estrictamente confidencial y no será usada para ningún otro propósito fuera de los de este estudio sin mi consentimiento. He sido informado de que puedo hacer preguntas sobre el proyecto en cualquier momento y que puedo retirarme del mismo cuando así lo decida, sin que esto acarree perjuicio alguno para mi persona.

Nombre del Participante

Firma del Participante

Fecha

(en letras de imprenta)

Appendix I: Conclusiones

La presente tesis doctoral ha intentado cumplir con una serie de objetivos con respecto a la disponibilidad léxica, la motivación en el aprendizaje de

una lengua extranjera, el género y el contexto de AICLE en dos lenguas extranjeras, el inglés y el francés. En concreto, ha pretendido explorar:

- las diferencias cuantitativas y cualitativas entre la disponibilidad léxica de los participantes en inglés y francés
- las diferencias cuantitativas entre la motivación hacia el aprendizaje del inglés en comparación con la motivación hacia el aprendizaje del francés
- las diferencias cuantitativas y cualitativas de la disponibilidad léxica en inglés y francés entre chicos y chicas
- la influencia de la instrucción AICLE en la disponibilidad léxica y en la motivación de los participantes en cada lengua

Para ello, se realizó un análisis de la disponibilidad léxica, la motivación, y el nivel de idioma en inglés y francés de alumnos de 9º, 10º y 11º grado (3º ESO, 4º ESO y 1º Bachillerato, respectivamente), quienes cursan asignaturas de instrucción AICLE en inglés y francés, a través de pruebas de disponibilidad léxica, cuestionarios sobre la motivación y pruebas C-test.

Los resultados de la primera pregunta de investigación demostraron que hay diferencias claras entre la disponibilidad léxica de los participantes en inglés y francés, con ventaja del inglés, probablemente debido a la competencia lingüística de los participantes en cada lengua. El análisis transversal también reveló la ventaja que tiene AICLE al exponer a los alumnos al vocabulario relacionado con el contenido de la clase, por ejemplo, los alumnos de 10º grado, que habían empezado a cursar economía en inglés, produjeron un nivel más alto de palabras en el centro de interés *Economía y Dinero* en ambas lenguas que los alumnos de 9º grado. Por otra parte, el análisis longitudinal reveló una diferencia aún mayor en la disponibilidad léxica en francés que en inglés desde 10º a 11º, quizás debido al nivel de partida más bajo que tenían los participantes en francés. Por otro lado, el análisis cualitativo indicó una gran similitud entre las dos lenguas. Sin embargo, se ha comprobado que existen diferencias claras en el ranking de los centros de interés más y menos productivos, lo que parece depender del idioma. Mientras los centros de interés relacionados con contenido en inglés fueron los menos productivos en los tres cursos en la prueba de disponibilidad léxica en inglés, el centro de interés relacionado con contenido en francés fue el más productivo en 9º y 11º grado en la prueba de

disponibilidad léxica en francés. Por consiguiente, parece que la exposición al vocabulario en el AICLE en francés marcó una diferencia más importante que la exposición al vocabulario en el AICLE en inglés. Estos hallazgos aportan una contribución importante al campo de la disponibilidad léxica, primero por responder a la llamada de Canga Alonso (2017) y destacar que sí hay una relación entre la disponibilidad léxica y el nivel de competencia lingüística, y segundo por demostrar la variación entre dos lenguas extranjeras diferentes, como el inglés y el francés, un asunto apenas abordado en investigación previa.

En cuanto a la segunda pregunta de investigación, encontramos una diferencia evidente entre la motivación hacia el inglés y el francés, ya que en todos los casos los participantes indicaron un nivel más alto de motivación hacia el inglés. Esa diferencia permanece constante durante los tres cursos, con poca variación entre los alumnos de 9º grado y 10º grado y entre los alumnos de 10º grado y 11º grado. Sin embargo, también encontramos actitudes más negativas hacia el aprendizaje del inglés en 11º grado, lo que está potencialmente relacionado con el hecho de que los alumnos generalmente ya no cursaron AICLE en este grado. En el análisis de la relación entre la disponibilidad léxica, motivación y nivel de idioma, observamos diferencias claras en cada idioma, ya que parece que existe una relación mucho más fuerte entre la disponibilidad léxica y la motivación y entre el nivel de idioma y la motivación en inglés que en francés. Estos resultados contribuyen a la investigación de motivación hacia el inglés y hacia otras lenguas extranjeras, y responden a la llamada de Oakes y Howard (2019) de investigar la motivación hacia las lenguas aparte del inglés y de integrar esta investigación con la del inglés para destacar las diferencias entre dos lenguas metas.

La tercera pregunta de investigación nos permite concluir que los participantes produjeron más palabras en inglés que en francés, independientemente del género, y que las chicas produjeron casi siempre más palabras que los chicos. Así vemos que la ventaja del inglés permanece para los chicos y las chicas por igual, mientras las chicas generalmente superan a los chicos en ambas lenguas. En cuanto a cada centro de interés, observamos la influencia de AICLE, por ejemplo, no había diferencias de género en el centro de interés *Economía y Dinero* en 11º grado, cuando los varones siguieron estudiando la economía a través del inglés, mientras las mujeres la estudiaron en español. También observamos que, aunque tanto los chicos como las chicas mejoraron de un curso

al siguiente, lo hicieron de manera distinta. El análisis cualitativo demostró pocas diferencias entre la disponibilidad léxica de los chicos y las chicas en inglés y francés. Sin embargo, los resultados indicaron que, aunque los chicos produjeron menos palabras que las chicas, puede que demuestren una mayor sofisticación léxica en función de su elevada proporción de palabras no compartidas y su mayor porcentaje de palabras fuera de lista. Estos resultados contribuyen a la investigación de disponibilidad léxica y género, aportando en particular observaciones interesantes sobre la manera distinta en que los chicos y las chicas progresan de un curso a otro, el efecto de AICLE, y las posibles diferencias cuantitativas y cualitativas. En cuanto a la motivación, podemos concluir otra vez que los participantes están más motivados hacia el inglés que el francés, independientemente de género, y que mientras los chicos y las chicas indicaron un nivel de motivación parecido hacia el inglés, había más diferencias en su motivación hacia el francés, sobre todo en el 10º curso. Revelamos también indicaciones de que los varones demuestran un nivel de motivación más alto hacia las asignaturas de AICLE que las mujeres en ambas lenguas. Estos hallazgos suponen una contribución importante a los estudios de género, ya que apoyan la sugerencia de Heras y Lasagabaster (2015) de que el contexto de aprendizaje afecta la motivación de los chicos y la de las chicas de manera distinta.

Por último, la cuarta pregunta de investigación ofrece una perspectiva interesante del efecto de AICLE, en concreto con respecto a la asignatura economía, que se relacionó con el centro de interés *Economía y Dinero*. De suma importancia es el hecho de que parece que la instrucción AICLE permite a los alumnos mejorar su disponibilidad léxica en la medida en que pueden efectivamente romper las distancias entre ellos y otros alumnos que no estudian este vocabulario relacionado con el contenido de la clase en la lengua meta. Es así aun cuando los otros alumnos tienen un nivel de idioma más alto y consiguen mejores resultados en otros centros de interés. Además, resulta que la motivación no tiene que ver con estas diferencias. En cuanto a las diferencias cualitativas, resulta que la instrucción AICLE conduce a más variedad en la primera palabra producida y a la producción de menos palabras al nivel K1-K5 y de más palabras fuera de lista en el centro de interés *Economía y Dinero*, lo que indica una mayor sofisticación léxica. En cuanto a la comparación de disponibilidad léxica en inglés y francés cuando un centro de interés está relacionado con una asignatura que se cursa en inglés o francés, concluimos

que los participantes producen más palabras en inglés, a pesar de que el centro de interés está relacionado con una clase en inglés o francés. Este es probablemente debido a la gran diferencia en los niveles de competencia lingüística, ya que el nivel de francés de los participantes fue considerablemente más bajo que su nivel de inglés. Sin embargo, hay indicación de que la lengua de AICLE sí afecta a la disponibilidad léxica: no se observó una diferencia entre los alumnos de 10º grado y 11º grado en inglés, pero se encontró una diferencia entre ellos en el único centro de interés relacionado con el francés, *Medio Ambiente y Clima*. Por consiguiente, aunque la diferencia en niveles puede ocultar una ventaja, parece que la instrucción AICLE ayuda a mejorar la disponibilidad léxica en el idioma de instrucción en centros de interés relacionados con el contenido de la clase. Respecto al análisis cualitativo, parece que el idioma de instrucción no tiene un fuerte impacto en los centros de interés en inglés y francés. Los anteriores resultados son de suma importancia para los estudios de AICLE, ya que responden a la llamada de Canga Alonso (2017) de explorar los centros de interés que están relacionados con las asignaturas AICLE, mostrando la ventaja de este contexto de aprendizaje en cuanto a la adquisición de vocabulario relacionado con el contenido de la clase. Los resultados también indican el efecto importante del nivel de idioma, lo que evidentemente hay que considerar cuando comparamos AICLE en inglés y en otras lenguas. Así respondemos a las llamadas de contrastar distintas lenguas en este contexto (Dalton-Puffer, Nikula y Smit, 2010; Cenoz et al, 2014; Merino y Lasagabaster, 2018a).

Todos los antedichos resultados benefician tanto a los investigadores como a los profesores. Desde la perspectiva de la investigación, este estudio ha respondido a varios llamamientos en investigación previa y ha llenado las lagunas que se han destacado, para así expandir nuestro entendimiento de estos asuntos intrincados. Además, los docentes pueden beneficiarse de estos hallazgos, que destacan consideraciones claves en el contexto multilingüe de la instrucción AICLE. En este contexto, donde una falta de preparación por parte de los profesores ha sido indicada repetidamente como problema, los resultados de este estudio pueden proporcionar a los interesados la información necesaria para prepararse mejor y proveer un contexto educativo más beneficioso.