

CORPUS-BASED CROSS-LINGUISTIC RESEARCH: DIRECTIONS AND APPLICATIONS [JAMES' INTERLINGUAL LINGUISTICS REVISITED]

MARLÉN IZQUIERDO FERNÁNDEZⁱ
UNIVERSITY OF LEÓN

Abstract: *This paper outlines the various possibilities offered by corpus linguistics for cross-linguistic research. Three main areas belonging to the so-called 'interlingual linguistics' field, namely Contrastive Analysis, Error Analysis and Translation Studies, are approached. The paper comments on the methodology followed to achieve their purposes, the direction of the research and the applications pursued, all of them being interdependent. This study also attempts at showing how the three areas of research contribute to the development of the field of cross-linguistic research by resorting to a common methodology, which makes them compatible with one another, even though their goals may be different. This methodology consists of corpus linguistics.*

Keywords: cross-linguistic research, contrastive analysis, translation studies, interlanguage, corpus linguistics, applications.

Resumen: *Este artículo revisa algunas de las posibles líneas de investigación que la lingüística de corpus ofrece al campo de los estudios interlingüísticos. Se abordan tres áreas representativas de este campo de estudio: el Análisis Contrastivo, el Análisis de Errores y los Estudios de Traducción. Se comenta la metodología adoptada por los investigadores en estas disciplinas para alcanzar sus objetivos, así como las diversas direcciones en la investigación y las aplicaciones que buscan realizar, siendo éstos tres factores interdependientes que se complementan mutuamente. Se pretende demostrar también cómo las tres áreas de estudio (AC, AE, ET) contribuyen conjuntamente al desarrollo de la investigación interlingüística a través de una metodología común, la lingüística de corpus. La Lingüística de Corpus se presenta como una metodología que hace posible el acercamiento entre disciplinas, aúna criterios y hace compatibles los resultados entre fronteras interdisciplinarias, aún cuando los objetivos de la investigación no sean los mismos.*

Palabras clave: investigación interlingüística, análisis contrastivo, estudios de traducción, interlengua, lingüística de corpus, aplicaciones.

1. Introduction

Scholars interested in the study of language have always resorted to language contrast and comparison, already before linguistics established as a science in its own right. Once linguistics takes off as a scientific discipline, an area of study seriously concerned with language contrast is born, namely Contrastive Analysis (hereinafter CA), at present much widely referred to as Contrastive Linguisticsⁱⁱ. Before CA appeared and during its development, language comparison and/or contrast proved to be fairly useful in broadening our knowledge on individual languages as well as on the relation between them. We have here an area of interlinguistic or cross-linguistic research, which, in its very beginning, unfolded within the applied branch of linguistics mainly committed to language teaching and learning.

The history of CA is shaped by ups and downs, the latter mainly due to the impossibility for the underlying theories to serve the linguistic needs of the time. However, from the 1980s onwards CA seems to emerge again, James' work playing an important role in its revival. By the time James published *Contrastive Analysis*, another area of cross-linguistic research had just started to develop as an autonomous discipline. We refer to the field of Translation Studies, as Holmes had named the various approaches to the study of translation and translations. However,

it was still early to talk about a discipline as such, while CA was struggling to recover its place within linguistic research. As a matter of fact, James considered that both areas along with so-called Error Analysis belonged to «a branch of linguistics, which I shall call ‘Interlanguage Study’, which is [...] interested in the *emergence* of these languages rather than in the finished product» (James 1980:3). Concerning the relation between these areas of study, James states that:

There are thus *three* branches of two-valued (2 languages are involved) interlingual linguistics: *translation theory* – which is concerned with the processes of text conversion; *error analysis*; and *contrastive analysis* –these last two having as the object of enquiry the means whereby a monolingual learns to be bilingual. (1980:4).

This classification by James has two important implications: 1) there exists within linguistics a niche concerned with the study of at least two different languages in contrast and/or in contact, here conceived of as a field of cross-linguistic research, and 2) these areas of study have their own objectives, and therefore evolve independently although there may be a meeting point, especially to contribute to the development of their domain of research and the linguistic science as a whole.

We do not totally agree with James on such a clear-cut disciplinary classification, as we will explain, but we take it as a representative domain of cross-linguistic research worth revisiting. Indeed, there are many other areas that would fit within interlinguistic research, such as code-switching, pidgins and creoles, dialectology, language typology, etc. We focus on these three branches because of the extensive traditional research done so far and also because we believe it is necessary to revisit James’ original theory in response to the current situation, which reflects the evolution and progress cross-linguistic research has undertaken over this time.

2. James’ *Interlingual Linguistics*: a niche of cross-linguistic research

Whereas James places translation theory apart, contrastive analysis and error analysis are considered to serve a common purpose, which is basically language teaching. This posture is still a reminiscence of the early days of CA, when it was believed that, by means of contrasting different languages, it would be possible to ‘detect’ the potential errors language learners would make, mainly considering the so-called *negative transfer* (Fries 1945), or negative influence from the first language which takes place in the production of utterances in the L2. Such a ‘contrastive-based’ methodology adopted by FLT extended successfully for some time, with Error Analysis emerging as a discipline or rather subdiscipline of CA interested in the *interlanguage*, i.e. that system between the mother tongue and the foreign language that is created in the learner’s mind. In the belief that it is in this *interlanguage* where errors take place, scholars decided to describe the languages independently, prior to contrasting them, and study how learners acquire their mother tongue and foreign language so as to detect problem-areas. So by that time, research within CA was mainly directed towards error analysis, and the results obtained were applied in language teaching. An obvious question is obliged here. If aware of the fact that errors were to be found in the interlanguage, why would they look for them in the process of language learning, even less in the description and contrast of the involved languages in their natural abode? Would it not have been more relevant to study instances of *interlanguage* itself? In other words, would it not have been more productive to consider Error Analysis as a discipline with its own object of study, goals and methodology to achieve them? That is, some kind of Contrastive Interlanguage Analysis? With the decline of CA in the 1970s, Error Analysis lost popularity as well. In fact, there is not so much current literature on Error Analysis as it was originally conceived, along with the fact that it is not given this name so frequently. Some new proposals are the study of learner language, interlanguage analysis, etc.

After the revival of CA enhanced by the new trends in linguistic research and methodology, a new branch still concerned with describing the process of language learning emerged. Such a branch has evolved parallel to CA but towards its autonomy, making it

possible for researchers in CA to broaden the scope of their research. CA is no longer a branch of applied linguistics exclusively focused on pedagogy. The array of applications of CA nowadays is much wider, expanding from a detailed description of languages in contact and/or in contrast, to facilitating applications outside linguistics proper, such as on the field of Translation Studies (from now on TS), which clearly exemplifies a cross-linguistic situation. The relationship between CA and TS has always been a matter of discussion. Halliday, for example, included TS within the domain of contrastive linguistics by assuming that

the theory and method for comparing the working of different languages is known either as 'comparative descriptive linguistics' or as 'contrastive linguistics'. Since translation can be regarded as a special case of this kind of comparison, comparative descriptive linguistics includes the theory of translation. (1965:112).

Following James, we prefer to talk about disciplines at the same level, but we disagree on any clear distinction between them because, as Rabadán states, «se trata de [...] campos cuyas fronteras son muy difusas» (1991:43). On the other hand, we see the connection that Halliday implies by considering that CA and TS may share a common methodology to attain their individual goals, which in turn implements the combination of results for an ultimate shared purpose, which is describing and improving intercultural communication. Bringing James' definition into consideration again, we cannot accept translation theory as just concerned with a mere process of text conversion any more. Throughout its development, TS has defined its object of study, which is no other than translation understood as both process and product. TS are to describe the working of language within an interlinguistic and therefore intercultural context, used to serve given social needs of communication. In other words, the process of translation demands the usage of at least two different linguistic codes. On the other hand, translation as product exemplifies such usage. It is mainly in the development of translation as process where CA serves as an auxiliary tool as Baker puts forward:

If translation is ever to become a profession in the full sense of the word, translators will need [...] acquire a sound knowledge of the raw material with which they work: to understand what language is and how it comes to function for their users. (Baker 1992:4).

In turn, «translated language is one of the ways in which the contrast of two languages is acted out. It is one of the ways in which linguistic contrast exists in social reality» (Mauranen 2002:165). Therefore, translation is also a way of knowing more about contrastive linguistics.

In short, CA and TS are acknowledged as two different disciplines with their own object, objectives and method of study, but they cannot ignore each other since both of them approach languages from a cross-linguistic point of view. Hence, they two necessarily have much to gain from each other as well as to contribute to the development of their common domain of study.

3. Corpus-based cross-linguistic research

At present, the applied nature of contrastive and translation research leads them in a parallel direction towards the improvement of intercultural communication, which is a converging point where results from the former are applied on the latter and vice versaⁱⁱⁱ. This is the current situation, result of the evolution of a context-based methodology for the study of cross-linguistic usage. Corpus linguistics has evolved as a common methodology for both CA and TS, making it possible to combine disciplines and applying the results obtained from CA to improving translator training as well as assessing translation quality. CA unveils interlinguistic correspondences which exemplify authentic and naturally-occurring language use in contrast. It is precisely the authenticity of these data which make them reliable enough to be used as translational norms which guarantee the acceptability of the translated text within the target context. At the same time, translation corpora have proven to be a reliable source of insights for CA since they reflect actual relations of equivalence between original texts and their translations. Out of the contrast between one and the other, it is now possible to extract data to be submitted to further contrast so as to evaluate translator performance and clarify assumed equivalent options between languages.

We argued before that the crisis CA faced in the 1970s derived in part from the fact that contrastive linguistic descriptions failed to respond to the pedagogic demands imposed by teachers, as well as from the difficulties in adjusting to the continuously changing linguistic theories (James 1980). The revival and following development of CA in the last two decades of the 20th century, notably in the 1990s, is associated with a linguistic theory, as well, but one whose focus of study is language use, that is, the functioning of language to serve social needs of communication, functioning which can only be accounted for if language is analysed in context. Such a linguistic theory is *functionalism* (Halliday 1985, Bondarko 1991, Chesterman 1998). Most importantly, this functional approach to language brought about a new methodology which made it possible to apply the results of contrastive analysis on areas where it had previously failed, as well as opening up its scope of application and bridging the gap between theory and practice. Such a methodology has considerably extended and it is nowadays the most reliable method in linguistic research, more specifically in applied linguistics, and even more precisely, in interlinguistic research. We are referring to Corpus Linguistics (from now on CL). CL approaches language in context, which is a conditioning factor of language use, hence the reason why it is the most suitable methodology for contrastive functional analysis^{iv}. CL has become a refined discipline capable of describing language used at different stages and/or for different purposes, be it during the stage of language learning, an act of communication among native monolingual speakers or an instance of translation, among other possibilities. We mention these three because they reflect the kind of language use that the branches of cross-linguistic research with which we are concerned here concentrate on.

In relation to the use of corpora in cross-linguistic research, Granger et al. consider that CL provides

a tremendous potential [...] in refining understanding of how languages relate to one another. It also demonstrates the value of using multiple corpus sources: the researcher who makes use of both comparable and translation corpora will gain a more comprehensive picture of whatever phenomenon is under investigation [...] corpus analysis of languages in contrast and in translation is a rich discovery process and what has been uncovered so far is only the tip of the iceberg. (Granger et al 2003: 12-13).

Granger's quotation above also hints at the important role that CL has played in bringing CA and TS together, which is clearly stated by arguing that:

Researchers in CL and TS have come to rely on corpora to verify, refine or clarify theories that hitherto had had little or no empirical support and to achieve a higher degree of descriptive adequacy. At the same time however, they do not always use exactly the same types of corpus and do not have the same research objectives. (Granger 2003:19).

CL offers different types of corpora depending on the area of study, and therefore the object of study and the objectives or applications. Within cross-linguistic research, we distinguish three major types of corpora, suitable for the areas outlined above.

Language learning is gaining ground within interlinguistic research, which is partly due to the insights provided by *learner corpora*. There are two major directions of language learning research by means of learner corpora: a) research on the learning process and usage of a foreign language and b) the learning process to become a translator. In relation to the first line of research we should mention the project carried out by Granger at the University of Louvain, for which the *International Corpus of Learner English* (ICLE) has been built (Granger [1993] 2003). This corpus contains texts written by students with a different linguistic background studying English as a foreign language. Researchers involved in this project adopt a contrastive methodology focusing on the interlanguage, which in turn may follow two different directions. One possibility implies contrasting native language usage with interlanguage, for example native English and English produced by French students. A second possibility contrasts the interlanguage produced by, for example, Swedish and French students of English. This approach is referred to as Contrastive Interlanguage Study.

Another project within this direction worth mentioning is the USE^v project developed at the Department of English at Uppsala University, which has created a learner corpus of up to

300 thousand words produced by Swedish students of English. Overall, we may consider this direction of research as a revisited and implemented version of the former Error Analysis. Nonetheless, its scope of study is slightly different and wider since it is not as much concentrated on predicting errors as on analysing which factors are more problematic as well as on the actual features of learning usage, which favours applications on language teaching, language learning assessment and course curriculum design.

Learner corpora have also proved to be a useful methodology to carry out and analyse translator training and performance. *Corpora created by translators (CCBT)* is a project led by Bowker at University of Ottawa (2003) through which the process of becoming a translator is assessed. By focusing on the usage of specific linguistic items (acronyms, false friends, etc.), researchers observe to what degree the learner's first language interferes in the students' performance, as well as the degree of target language acceptability attained. The applications of this direction within cross-linguistic research are twofold: first, it has direct applications on translator training, and second, the insights drawn may be useful for language teaching in general too.

Both translation and language teaching constitute the major areas of application of contrastive analysis today. Research within CA is almost by definition corpus-based, with CL establishing as the main methodology in the field, already since the mid 1970s as announced by Filipović, for whom «the first problem facing researchers engaged in a contrastive analysis project is that of the method to be adopted. Immediately after that comes the closely connected question of the corpus» (1974:51). Due to the object of study, linguistic theory and future applications in mind, researchers may opt to use different types of corpora for a CA project. *Comparable corpora* have been considered the most suitable corpus type in the belief that they provide contrastive linguists with authentic language data naturally occurring within native contexts. A useful direction within CA using comparable corpora consists of contrasting the realization of a semantic function or meaning in two (or more) different languages so as to retrieve interlinguistic correspondences between them (Labrador 2000, Ramón García 2002). The degree of correspondence is established in terms of frequency and patterns or co-texts of occurrence of the individual formal resources that each language has for conveying that common function. The outcome of this line of research does not only broaden our knowledge of the functioning of individual languages, but it also sheds light on the relation between them, which has important implications for interlinguistic communication. In fact, the correspondences detected and assumed to be functional or translation equivalents may well be applied on the field of TS as norms according to which we may either guarantee or assess linguistic acceptability within the target context. That is, corpus-based CA may have applications on translator training and translation quality assessment.

The reliability and adequacy of comparable corpora for CA lie on the researchers' need to measure the real equivalence between languages as they occur naturally, without being affected by other linguistic codes. This explains why some researchers have eagerly argued against the use of *translation corpora* in CA saying that:

Translations, however good and near-perfect they may be (but rarely are), cannot but give a distorted picture of the language they represent. Linguists should never rely on translations when they are describing the language. That is why translations have no place in reference corpora. Rather than representing the language they are written in, they give a mirror image of their SL (Teubert 1996:247).

Indeed, translations represent a distorted image of native language use, but this must not necessarily be a drawback for researchers. On the contrary, using translation corpora in CA may be the right way to proceed if the ultimate goal of our research is to have applications on translation. After all, translational corpora are the best repositories of translated language or translator usage of the language. Therefore, the more we know about it, the better the applications on the field of translation. *Parallel corpora*, that is, bilingual corpora composed of original texts in a given SL and their translations into a TL, are more and more frequently used in CA. A parallel corpus may give new insights between the languages compared that would not

be noticed by working with individual monolingual corpora (Johansson 2003) and, most importantly, parallel corpora reflect relations of sense as well as of form between languages (Mauranen 2002), the identification of which would turn out a daunting task if working with comparable corpora. Parallel corpora have important implications for CA since they open up a new direction of research whereby the starting point for linguistic contrast is a formal resource whose assumedly equivalent expressions are provided by the corpora. In other words, parallel corpora may well provide data for functional contrast even when studying formal resources. Aware of the interference of the SL in the translated texts, the insights provided by this type of corpora should be verified or compared with data from comparable corpora or a monolingual corpus of the target language, so as to analyse the translational options or equivalents identified in their natural context. The results obtained out of this kind of project may have the same applications as those provided by comparable-corpus-based CA. The direction of research is the opposite, from form-to-meaning instead of from meaning-to-form, but either study is framed within the functional approach because functional equivalence is the criterion for contrast (Bondarko 1991). Moreover, meaning and form are one and the same, without meaning there is no form and vice versa, meaning-form are connected through the function they perform together. Consequently, «means are shaped by ends» (Chesterman 1998:63) so that either way would take us to completing a functional contrast of the languages involved.

So far we have discussed corpus-based language learning and CA and their applications on translation. But TS has established as an autonomous discipline over the years, with its own methodology of analysis and applications, for which *translation corpora* play an obviously important role. We could mention as two of the most important directions of research within the field the following: a) the study of the so-called *universals of translation* and *translationese* and b) the *norms of translation*. Related to these lines of research, one of the major applications in TS is translation quality assessment (TQA), which in turn favours applications on translator training.

By analysing substantial amounts of translated language, researchers expect to find out patterns or regularities that would come to characterize translated language. The *universals of translation* (Baker 1993, Laviosa 1996) are considered features of all translated language irrespective of the SL, whereas *translationese* would belong with deviated uses of the target language due to the interference of the source language (Mauranen 2002). These features are data obtained when translation is approached as *product*. On the other hand, by *norms of translation* we imply here those recurrences of translator performance during the *process* of translation, in an attempt at fulfilling the expectations of the target context in terms of communicative acceptability as well as functional equivalence with the source text. When studying *norms of translation*, the target or recipient context of the translated text is a conditioning factor which requires a sound knowledge of the target linguistic code and its communicative function in due context. As such, corpus-based CA has again something to contribute to this line of research, since it may bring the two different contexts to actual contrast.

TQA would resort to the results obtained from analysing a great amount of translated language to find out all those target textual and linguistic features which may well hint at where, how, and possibly why, they depart from their source features or the expected target-context ones. That is, applying the results provided by corpora on translation assessment is a reliable and objective way of improving translation activity. Likewise, the descriptive nature of the insights provided by corpus-based studies favours their contribution to the writing of a theory of translation.

4. Conclusion

To sum up, by borrowing James' interlingual linguistics (James 1980), we have selected a representative domain of cross-linguistic research in order to show how corpus linguistics has brought areas of study *back to life*, shaping its new development, at the same time as it has

enhanced that of emerging ones. Cross-linguistic research is nowadays defined as 'corpus-based', with areas of study, long ago kept apart and subsidized, sharing a common methodology which serves their individual purposes but makes it possible to combine results and extend applications. This is a point of convergence where the field of cross-linguistic studies reaches its systematic whole. Functionalism takes also an important part in the stability and continuous growth of corpus-based cross-linguistic research. The functional approach has put forward a dynamic conception of language as a tool for communication, whose description is the viable way towards its own improvement.

Although the *what* and *how* of the field differ from those of the early 1980s, the *why* is not so different. Whilst James considered that those areas belonging to the field of *interlanguage studies* were concerned with the *emergence* of languages, the present-day cross-linguistic disciplines that we have revisited in this paper also seek to know more about that process of language emergence in two major contexts: learning to use language as a translator and learning to speak a second language, with the three of them – CA, TS & the now called CIA- having a part in the process.

Notes

ⁱ Departamento de Filología Moderna, Universidad de León.
Campus de Vegazana s/n 24071, León –Spain.

E-mail: dfinmif@unileon.es

ⁱⁱ In this paper we will use the term Contrastive Analysis (CA) instead of Contrastive Linguistics (CL) to avoid confusion between the latter and Corpus Linguistics (also CL). However, whenever we refer to CA we imply Contrastive Linguistics.

ⁱⁱⁱ As already stated, there are various possibilities of application but in this paper we are mainly concerned with the interdependence between CA and TS which is also reflected through the applications of one over the other.

^{iv} CL may be compatible with other linguistic theories, but it clearly meets the principles of functionalism, the most important of which is the study of linguistic usage in context so as to describe language.

^v Uppsala Student English Corpus (Axelsson & Berglund [1999] 2002)

References

- BAKER, MONA, «Corpus Linguistics and Translation Studies», in BAKER, M., FRANCIS, G. et al (eds.) *Text and Technology. Studies in Honour of John Sinclair*, 1993, pp. 233-250.
- BONDARKO, ALEXANDER, *Functional Grammar. A field Approach*, Amsterdam/Philadelphia, John Benjamins, 1991
- CHESTERMAN, ANDREW, *Contrastive Functional Analysis*, Amsterdam/Philadelphia, John Benjamins, 1998.
- FILIPOVIĆ, RUDOLPH, «The use of a corpus in Contrastive Studies», in RAABE, H. (ed.) *Trends in Kontrastiver Linguistik*, Tübingen, Gunter Narr, 1974, pp. 51-66.
- GRANGER, SYLVIANE, LEROT, J. & S. PETCH-TYSON, *Corpus-based Approaches to Contrastive Linguistics and Translation Studies*, Amsterdam/ Nueva York, Rodopi, 2003.
- HALLIDAY, M.A.K., *An Introduction to Functional Grammar* (2nd edition), London, Arnold, 1985
- HOLMES, J.S. «The Name and Nature of Translation Studies», in *Translated! Essays and Papers on Translation Studies*, Amsterdam, Rodopi, 1988, pp. 67-80.
- JAMES, CARL, *Contrastive Analysis*, Singapore, Longman, 1980.
- LABRADOR, BELÉN, *Estudio Contrastivo de la Cuantificación Inglés-Español*, León, Universidad de León, [2000]2005.
- LAVIOSA, SARA. «Comparable corpora: Towards a corpus linguistic methodology for the empirical study of translation», in HELEN M. & LEWANDOWSKA TOMASZCZYK

- B. (eds.). *Translation and Meaning, Part 3*, Maastricht, Hogeschool Maastricht, 1996, pp.153-163.
- MAURANEN, ANNA, «Will 'translationese' ruin a CA? » In *Languages in Contrast*, 2 (2), 2002, pp.161-185
- RABADÁN, ROSA, *Equivalencia y Traducción. Problemática de la equivalencia transléctica inglés-español*. León, Universidad de León, 1991.
- RAMÓN GARCÍA, NOELIA. *Estudio Contrastivo Inglés-Español de la caracterización de sustantivo*, León, Universidad de León, 2002.
- TEUBERT, WOLFGANG, «Comparable or Parallel Corpora? » In *International Journal of Corpus Linguistics*, 9 (3), 1996, pp. 238-64.