

A multidimensional analysis of two registers of English for Navy submariners

Yolanda Noguera-Díaz 

yolanda.noguera@upct.es

Universidad Politécnica de Cartagena, Spain

Noguera-Díaz, Y. (2022). A Multidimensional analysis of two registers of English for Navy submariners. *Language Value*, 15(2), 81-111. Universitat Jaume I ePress: Castelló, Spain.

<http://www.languagevalue.uji.es>.

December 2022

DOI: <https://doi.org/10.6035/languagev.7019>

ISSN 1989-7103

ABSTRACT

This research explores the differences and the similarities found in two corpora representative of two registers of relevance for Navy submariners in the Spanish Navy Submarine Warfare School. It shows cases in a range of analyses based on multi-dimensional analysis, characterizing these two submariner registers relative to Biber's 1988 dimensions of register variation. The findings can potentially inform professional language teaching in such contexts. It is argued that linguistic that can inform professional language teaching in such contexts. It is argued that linguistic variation among the texts affords the identification of both converging and diverging patterns of variation across dimensions of use.

Keywords: *Corpus linguistics; multidimensional analysis; linguistic variation; professional languages; English for the Military*

I. INTRODUCTION

Text-linguistic register analyses examine “the lexico-grammatical features that are frequent and pervasive in [...] texts that all share the same situational characteristics, and thus, all represent the same register” (Biber, 2019b, pp. 46-7). An instance of such analysis is multi-dimensional analysis (MD analysis), which has been widely used in the exploration of a range of professional and academic registers leading to an increased understanding of how the frequency and distribution of linguistic features contributes to variation.

Although pedagogical applications such as the development of educational materials have been suggested (Biber, 1998; p.236), there is a lack of research that has explored how MD analysis can potentially inform language teaching in professional, non-academic contexts (Friginal & Roberts, 2020). Our research examines how MD analysis can potentially inform applied linguists and language teachers’ choices of texts across discourse domains (Biber, 2019) when designing curricula for specialized languages. This study sets out to analyse a type of English for the Military known as Submarine English (SE) used by Navy submariners. This article uses MD analysis to reveal aspects of variation in two corpora representative of SE: (1) a corpus of professional magazines and (2) a corpus of manuals for maritime salvage and rescue of submarines. Thus, this research sets out to explore the linguistic characteristics of these two submariner registers relative to Biber’s 1988 dimensions of register variation. In doing so, this study discusses the potential contribution of quantitative text-linguistic studies of register variation (Biber, 2019a) to corpus-informed pedagogy of non-academic, professional languages.

This paper is structured as follows. Section II revises the contributions of MD analysis to the study of registers. Section II.2 describes the research methodology, while section III examines the results and discussion of our analysis. In Sections IV we provide some conclusions and offer some insights into the contribution of variation analysis to the pedagogy of specialized and professional registers.

II. LITERATURE REVIEW

II.1. Multidimensional analysis and the study of variation across registers

MD analysis seeks to interpret linguistic data in the light of language variation across registers. In the MD analysis tradition, a register is a variety of language associated with a particular situation of use that displays specific communicative purposes (Biber and Conrad, 2009, p. 6). Register analysis explores the link between use and a social situation with a view towards explanation. While register analysis looks at lexical phraseology, as well as grammatical and lexico-grammatical features of a text, the situational analysis comprises characteristics of texts such as the communicative purpose, mode, setting and participants.

MD analysis (Biber, 1988; Biber & Conrad, 2009; Biber, 2019a) identifies systematic patterns of variation across registers. Co-occurrence patterns are interpreted as dimensions of variation based on the shared communicative functions of the co-occurring features. Each dimension is associated with a set of linguistic features which tend to occur in texts. In the analysis of each dimension, we obtain sets of features, both positive and negative, that are in complementary distribution. If, for instance, a set of texts shows a high frequency of common nouns, it will also tend to have a high frequency of adjectives (Biber & Gray, 2013).

Biber's (1988) study identified five main linguistic dimensions of language use that have been widely used by researchers to identify variation in most types of texts. MD analysis enables a discourse domain to be described quantitatively and functionally (see section III). From a quantitative perspective, dimensions scores quantify the extent to which they use features associated with the dimension and at the same time it is based on frequencies of the co-occurring features. Functionally, dimensions are interpreted based on the analysis of shared functions of features, analysis of text excerpts and register distributions. In the following sections, we provide a breakdown of how MD analysis has been used to study register variation.

II.1.1. Strategies

ESP (English for Specific Purposes), and by extension languages for specific purposes (LSP), is an area of inquiry and practice either in workplaces (Hutchison and Waters, 1987) or in academia such as English for biology (Gray, 2013). Despite the underlying motivation to improve curricula and classroom practice, pedagogical applications of MD analysis have not received as much attention as the linguistic description of specialized corpora. Some domains have received, however, some substantial attention. Crosthwaite et al (2019) collected a corpus of dental public health papers which includes experimental research papers, Dentistry professional research reports and Dentistry case studies. Their MD analysis explored the linguistic features employed by Dentistry professionals and undergraduate students' writings. The analyses revealed the extensive use of the passive voice in professional writing as well as two dimensions of use involving a pervasive style (D4) and a more informative approach (D2).

Global aviation has similarly received some attention. Friginal and Roberts (2020) compared the functional features of linguistic dimensions in six spoken corpora: call centers, exploration aviation, maritime English, home calls, switchboard and general American conversations. They used the linguistic dimensions in Friginal (2009): Dimension 1 (Polite and elaborated information vs simplified narrative), Dimension 2 (Planned talk) and Dimension 3 (Managed Information flow). In Dimension 1, Call Centre language showed the highest scores due to the number of polite markers (e.g., please, thank you) whereas Aviation language showed the lowest score. Exploration Aviation and Maritime English corpora yielded very similar scores in Dimension 2 (Planned talk), which highlights the fact that procedural and instructional instances are common in both registers.

The analysis of linguistic variation using MD analysis has gained some traction (Friginal & Roberts, 2020; Ren & Lu, 2021). However, little is known about the adaptation of such findings to the teaching of professional and specialized languages, particularly in non-university contexts. Due to the dearth of teaching materials for the military (Noguera-Díaz, 2019), the present study examines a corpus of submarine English (SE) applying text linguistic register analysis that goes beyond previous efforts focused on

the examination of discrete features such as noun phrase complexity (Author, 2020). This paper examines two registers of relevance to the students in the Navy school: professional submarine magazines and salvage and rescue manuals and technical reports. This research addresses the following research question:

1. What are the linguistic characteristics of the two submariner registers analysed relative to Biber's 1988 dimensions of register variation?

III. METHODOLOGY

In this section, we discuss the corpora investigated, the methods that were adopted in the MD analysis as well as the statistical test used.

III. 1. Corpora

As Noguera-Díaz (2019, 2020) noted, access to classified texts for instructional purposes is restricted to the military on-site and the analysis of classified sources is not possible. Accordingly, the choice of the texts for the subject English for Navy Submariners was determined by the management of the Navy Submarine Warfare School (NSWS). Our research examines (1) a corpus of professional military submarine magazines (CMSC) and (2) a corpus of manuals used in submarine search and rescue (SAR). CMSC and SAR represent two registers (professional magazines and manuals) that are relevant for the training and language learning of the Navy submariners at the NSWS.

CMSC (Noguera-Díaz, 2019) is a corpus of US military magazine articles curated by the Spanish Ministry of Defence and distributed in printed form. Each issue is made up of selected articles from a pool of fourteen specialized magazines. The CMSC is made up of 822,755 words and comprises 12 years of curated texts published across a wide range of different professional magazines regularly read by trainees and used by language instructors for language learning purposes. It contains a total of 952 different

texts. For the sake of our analysis, each year issue (n=36) has been computed and analysed separately. The SAR corpus is a collection of 16 non-classified manuals and reports recommended by the NSW. They are used as references for a compulsory subject on maritime search and rescue. These texts have been selected and read by professional trainees and used by tactical and language instructors. The SAR corpus is made up of 717446 words and comprises texts published by either professional associations such as the NATO Standardization Agency (NSA) or publishers such as Defence Research and Development Department, Canada. Some of the manuals are published by organizations based on countries where English is not an L1 language. Each manual (n=16) has been computed and analysed separately. Further details about the CMSC and the SAR corpora can be found online in Appendix 1 (https://www.researchgate.net/publication/365835753_CMSC_AND_SARdocx).

III. 2. Corpus analysis: multidimensional analysis

The two corpora were POS tagged and analysed using MD analysis (Biber, 1988; Biber, 2019a). MD analysis “empirically analyses the ways in which linguistic features co-occur in texts and the ways in which registers vary with respect to those co-occurrence patterns” (Biber, 2019b, p. 49). The five dimensions of language use in Biber (1988) were computed and a factor score was calculated for each of them using the multi-Dimensional tagger (MAT) (Nini, 2019). This has been described as the 1988 model of variation (Biber, 2019b). Each dimension of use has distinct functional underpinnings. All frequencies of the linguistics features analysed are standardized to a mean of 0.0 and a standard deviation of 1.0 before the computation of the factor score. A factor score is a numerical value that indicates a text relative standing on a latent factor in factor analysis. According to Nini (2019, pp. 9-17), the dimension scores produced by MAT are reliable as shown by the replication of Biber’s (1988) analysis of English language corpora.

II.2.1. Hypothesis testing

We used the Kruskal-Wallis H test to determine statistically significant differences between two or more groups of a dependent variable, in our case the five dimensions. The Kruskal-Wallis H test is a rank-based nonparametric test that can be used to test whether Dimension scores are different between CMSC, and SAR corpora based on the use of mean ranks. To know where any differences lie, post hoc tests were run.

IV. RESULTS

The following sections show the overall dimension scores for each of the two corpora (IV.1) and a discussion (IV.2) of the main dimensions of variation following Biber (1988). We will pay special attention to the dimensions showing statistically significant differences and will showcase excerpts where some of the most relevant linguistic features are found in the two corpora analysed. The samples used below showcase texts that offer high degrees of either inter-corpus or intra-corpus linguistic variation. Readers are invited to interpret individual text scores, provided in brackets, against the backdrop of the corpora Dimension scores. Section IV.II offers a summary of the results.

IV. 1. Corpus dimension scores

Table 1 shows the mean scores of the CMSC and the SAR corpora for Biber's (1988) dimensions 1-5.

Table 1. *Dimension scores of the CMSC and the SAR corpora*

Dimension	Dimension interpretation (Biber, 1998)	CMSC	SAR	Statistically significant differences?
		Mean score	Mean score	
1	Involved vs informational production	-19.95	-19.56	No
2	Narrative vs non-narrative concerns	-2.9	-4.59	Yes
3	Explicit vs situation dependent reference	5.77	8.12	Yes
4	Overt expression of persuasion	-1.27	0.95	Yes

5	Abstract vs non-abstract information	2.47	1.64	No
---	--------------------------------------	------	------	----

Statistically significant differences between the two corpora were found for Dimensions 2, 3 and 4. In the following paragraphs, results of the MD analysis of every component of the two corpora are provided. Results for the individual components of the corpora are shown chronologically (i.e., 2000 to 2012) in the following figures.

IV.1.1. Dimension 1: Information production orientation

As shown in Figure 1, both corpora show a similar Dimension 1 (D1) score. The score range fluctuates between CMSC3 (-15.18) and CMSC34 (-24.23), and between SAR10 (-15.68) and SAR7 (-28.73). Both corpora show a marked information orientation with a low impact of interpersonal features. Figure 1 shows D1 scores of the corpora analysed.

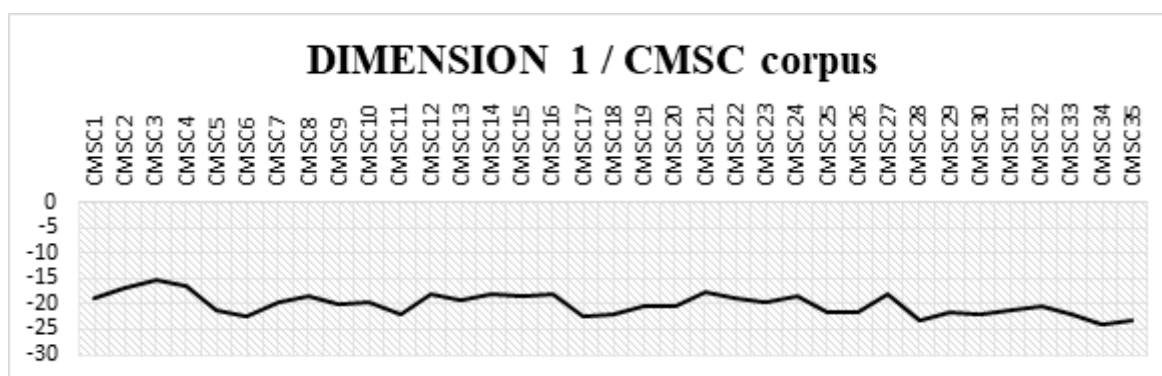


Figure 1. Dimension 1 scores for the CMSC and SAR corpora

Professional magazines (CMSC) and manuals and reports (SAR) show scores below the means for academic prose (-15) and official documents (-18) in Biber (1988). The distribution of D1 scores was similar for the two corpora, as assessed by visual inspection of a boxplot. A Kruskal-Wallis test was conducted to determine if there were differences between the two corpora. D1 scores were not significantly different between the two corpora, $H(3) = 1.076, p = .201$.

Both corpora display a high frequency of features associated with informational production, such as nouns, attributive adjectives, long words, prepositions, type/token ratio, agentless passives, place adverbials and past participle postnominal clauses (Biber, 1988). SAR texts, however, show a higher mean score for nouns (34.8), nominalizations (5.2) and agentless passives (1.5) than CMSC texts. These features suggest that the texts in the SAR may show a marked “informational focus and a careful integration of information in a text” (Conrad and Biber, 2001:24). It is important to bear in mind that nouns are the principal way employed by writers to refer to concepts or entities (Conrad, 2001) and are essential to display dense information packaging (Biber et al., 1999).

Sample 1 includes CMSC3 and CMSC34 texts. CMSC3 includes the article *Dog fighting submarines* which was written by a North American Captain for the *Submarine Review Journal*. He writes about his past experiences as a submariner and as an expertise on nuclear submarines and technical innovations. CMSC34 includes the article *Canadian sub overhaul begins with Chicoutimi* was published in *Jane’s Defence Weekly*. Here the journalist discusses submarine in-service contracts and capabilities.

Sample 1: CMSC CORPUS

Nouns in bold and attributive adjectives underlined

CMSC3 (2000.3) = (-15.18)

With a powerful new passive sonar and enormous mobility.

CMSC34 (2011.3&4) = (-24.23)

*The diesel-electric boat is being overhauled at Victoria **Shipyard’s** covered **facility** in British Columbia...*

Sample 2 includes SAR7 and SAR10 texts. SAR7, *An Assessment of the CF Submarine watch schedule variants for impact on modelled crew performance*, was published by the *Canadian Defence Department*. This technical report was produced after a fire on board of a Canadian submarine. SAR10, ATP-57_B, is a NATO non-confidential Allied Tactical Procedures publication which describes some basic concepts related to

Command, control and communications during the rescue operations, mainly how Submarine Rescue Operations or exercises should be conducted. SAR07 and SAR10 contain the highest mean scores for nouns in both corpora.

SAMPLE 2: SAR CORPUS

SAR7 (2009): (-28.73)

Nouns in bold and attributive adjectives underlined.

The primary **objective** of this field **study** was to evaluate whether enhanced white light would promote circadian entrainment.

SAR10 (2011) = (-15.68)

SMERAT personnel require to know what the **capabilities** of individual SPAG teams are and how to interact with them. There is inter-country **variability** between the **SPAG teams**.

IV.1.2. Dimension 2: non-narrative orientation

While professional magazines (CMSC) and manuals and reports (SAR) share a non-narrative orientation, both corpora show different dimension scores (see Table 1). Dimension 2 (D2) score range varies between CMSC21 (-0.1) and CMSC23 (-3.98), and between SAR7 (0.22) and SAR16 (-5.8), respectively. Figures 2 and 3 show D2 mean scores of the corpora analysed.

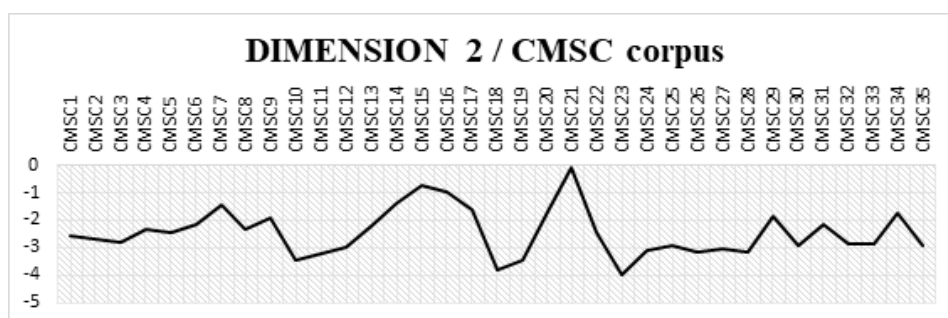


Figure 2. Dimension 2 scores for the CMSC corpus

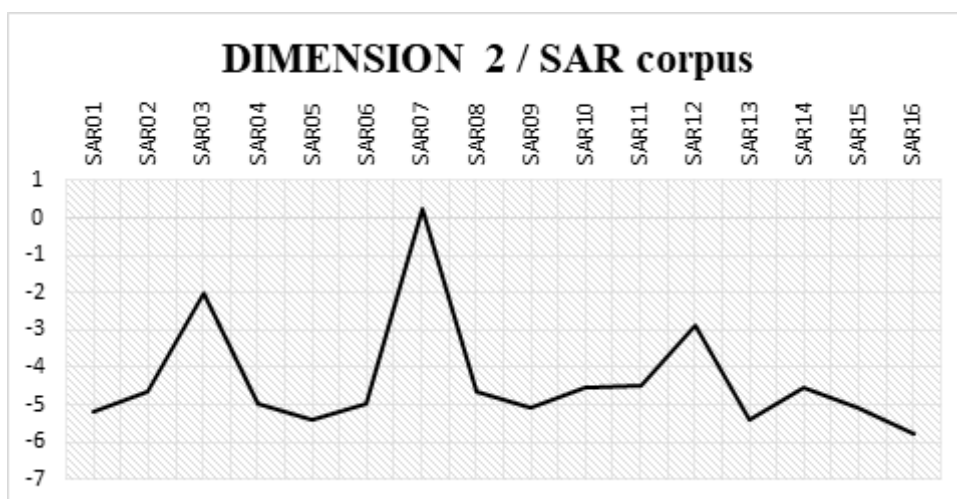


Figure 3. Dimension 2 scores for the SAR corpus

CMSC scores are relatively closer to narrative concerns (-2.9) in contrast to SAR texts (-4.59). Pairwise comparisons were performed using Dunn's (1964) procedure with a Bonferroni correction for multiple comparisons revealed statistically significant differences in D2 scores between the CMSC and SAR corpora $H(2) = -3.991, p = .0005$. While the CMSC corpus shows a mean score (-2.9) like that of hobbies and broadcasts texts (-3), the SAR corpus displays a mean score (-4.59) well below these two registers according to Biber (1988).

Registers with high negative D2 scores include professional letters, academic writing, and official documents. The CMSC corpus shows a mean frequency of 8.56 attributive adjectives per 1,000 words whereas the SAR texts show a lower mean (6.2). In addition, present tense verbs display a higher mean in the CMSC corpus (4.37) versus the SAR corpus (3.12).

Biber's (1988) original MD analysis shows that the features with positive weights in D2 are past tenses, third person pronouns, perfect aspect, public verbs, synthetic negation, and present participial clauses. The features with negative weights are present tense verbs, attributive adjectives, and past participles with deletion. CMSC texts show a higher mean for past tenses (1.79), third person pronouns (0.53) and perfect aspect tenses (0.55) whereas the mean of the SAR texts is (0.59) for past

tenses, (0.32) for third person pronoun and (0.18) for perfect aspect tenses. The excerpts below illustrate the range of variation found across the two corpora. Let us take CMSC21 (-0.1) and CMSC23 (-3.98), and SAR07 (0.22) and SAR16 (-5.8) as instances of extreme variation in the data.

Sample 3 includes CMSC21 and CMSC23 texts. CMSC21 (2006.4) includes *The value of Submarines* published in the Military Technology journal the last term of 2006. This article discusses the economic benefits for Texas provided by the submarine industry. CMSC23 (2007.3) includes *the Modernization of Chilean Navy*, published in Naval Forces. The article discusses the increment of the Chilean military budget.

Sample 3: CMSC CORPUS

CMSC21 (2006.4) = (-0.1)

Perfect verb tenses are underlined and **public verbs** in bold.

The region has supported these activities reflexively and often half-heartedly.

*The Navy **claims** to need at the present rate of building one submarine a year.*

CMSC 23-(2007.3) = (-3.98)

Attributive adjectives in bold.

*The country is situated on the most **peaceful** continents in the world, and enjoys fairly **good** relations with all nations of the region.*

CMSC 21 displays the highest positive scores in perfect aspect tenses (0.56) and public verbs (1.39), illustrated in the sample above, which explains why the mean score of this text on D2 (-0.1) is close to a narrative register. On the contrary, CMSC23, shows the highest score of attributive verbs (9.63), together with a range of past participles with deletion (0.38), which exhibits the high negative score for D2.

Sample 4 includes SAR7 and SAR16 texts. SAR7 was discussed previously. SAR16 is a volume entitled CORPAS-SARSAT: *Search and Rescue Satellite Aided Tracking*. It was written by a steering Corpas-Sarsat committee and published in 2019.

Sample 4: SAR CORPUS

SAR7 (2009) = (0.22)

Past tenses are underlined and **public verbs** in bold.

*The entire subject population **reported** sleepiness in the middle of the scale thus confirming that they were quite sleepy.*

SAR16 (2019) = (-5.8)

Present simple verbs underlined and **past participles** with deletion of relative in bold.

*This document contains the minimum requirements that apply to Cospas-Sarsat distress **beacons**.*

*Beacons type **approved** by Cospas-Sarsat for operation at 406.025 MHz*

SAR corpus shows that highest scores within this dimension (SAR7, 0.22) as well as the lowest negative score of this corpus (SAR16, -5.8) for Dimension 2. SAR7 shows a higher positive score for verbs in past simple (3.56) and public verbs (0.50) as seen above. While SAR7 shows some narrative orientation, SAR 16 shows the minimum score mean of both corpora in D2.

IV.1.3. Dimension 3: textual elaboration

In Dimension 3, professional magazines (CMSC) and manuals and reports (SAR) texts show very different mean scores: 5.77 and 8.12, respectively (Table 1). However, both CMSC and SAR share a clear orientation towards context independence and textual elaboration (Biber, 1988). Dimension 3 (D3) score range varies between CMSC19 (7.08) and CMSC21 (-3.69), and between SAR05 (15.59) and SAR15 (4.28). Figures 4 and 5 show D3 mean scores of the corpora analysed.

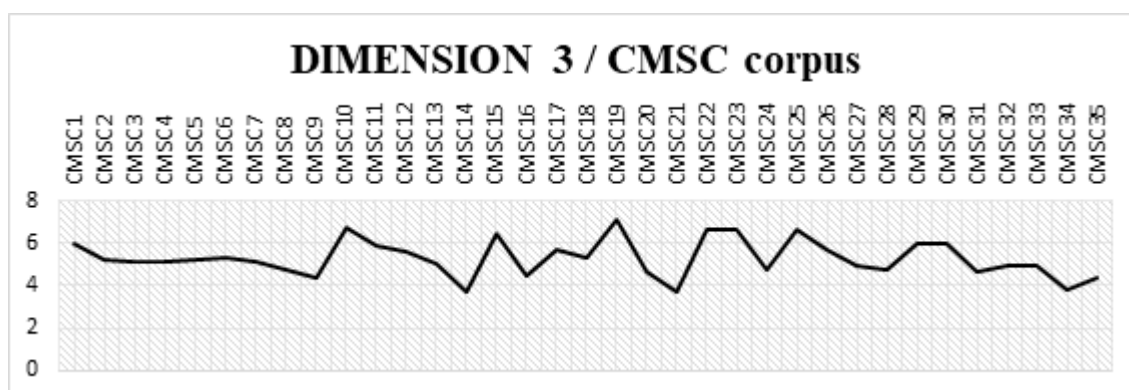


Figure 4. Dimension 3 scores for the CMSC corpus

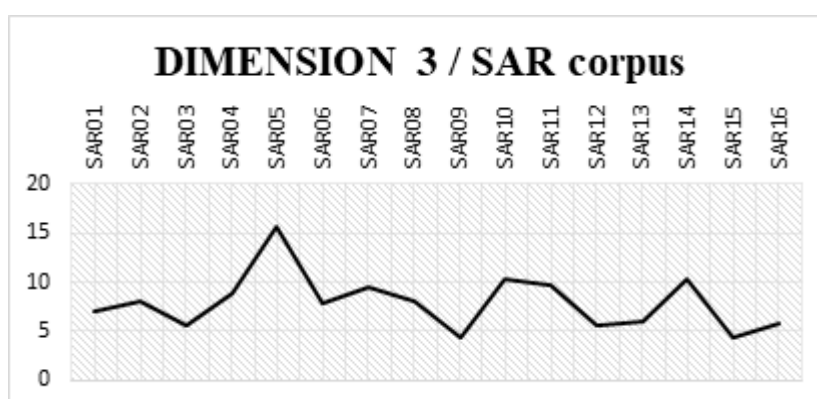


Figure 5. Dimension 3 scores for the SAR corpus

CMSC shows a mean score (5.77) above that of academic texts (4.2) in Biber (1988), while the SAR corpus shows a mean score (8.12) above official documents (7.3). Pairwise comparisons revealed statistically significant differences in dimension 3 scores $H(2) = 2.829$, $p = .005$. High positive scores in this dimension show independence from context, whereas low scores display dependence on the context. Linguistic features with a positive weight in D3 include wh-relative clauses in object position, pied-piping relatives (preposition + a relativizer), wh-relative clauses in subject position, phrasal coordination and nominalization (Biber, 1988). Linguistic features with negative weights on D3 include time and place adverbials.

The SAR corpus in D3 shows a higher means score for nominalizations (5.25) than CMSC texts (3.61). However, the mean frequency for phrasal coordination (1.17) is

identical in both corpora, as well as the wh-object relative clauses on object position (0.01). Relative clauses are relatively infrequent in the two corpora. Time and place adverbials are more frequent in CMSC texts 0.36 and 0.31 versus 0.14 and 0.21 in SAR texts, respectively.

Sample 5 includes CMSC21 and CMSC19 texts. CMSC21 (2006.4) includes *Warfare: capabilities and assets required*; an article published in 2006 in the Naval Forces magazine. CMSC19 includes *Iran tests high-speed*, originally published in 2006 in Undersea Enterprise News.

Sample 5: CMSC CORPUS

CMSC21(2006.4) = (-3,69)

Place and time adverbials in **bold**.

*Spending 49 days **at sea** the boat arrived **in Simon's Town**, **some 45 kilometers southeast of Cape Town** after a voyage of **6,600 nautical miles**.*

CMSC19 (2006.2) = (7.09)

Underlined phrasal coordination

The United States and its Western allies have been watching Iran's progress in missile capabilities with concern.

CMSC19 texts have the highest positive mean value (7.09) in the corpus, with higher scores in phrasal coordination (1.30) and nominalization (4.09). CMSC21 has the lowest mean value in D3 (-3.69), and the highest scores for time adverbials (0.40) and place adverbials (0.42).

Sample 6 includes SAR15 and SAR5. SAR15, *Specifications for CORPAS and SARSAT*, is a technical document that explains the requirements for the development of 406 MHz maritime distress beacons, Emergency Position-Indicating Radio Beacons (EPIRBs) and Personal Locator Beacons (PLBs) for personal use. SAR 5, IAMSAR V.1, stands for *International aeronautical and maritime search and rescue manual* discusses common aviation and maritime procedures to provide Salvage and Rescue Services following

the International Convention for the Safety of Life at Sea (SOLAS). It was published jointly by the International Maritime Organization (IMO) and the International Civil Aviation Organization (ICAO).

Sample 6: SAR CORPUS

SAR 15 (2019) = (4.28)

Nominalizations in bold.

*The beacon shall commence **transmissions** upon **activation** even if no valid position data are available.*

SAR 5 (2005) = (15.59)

Nominalizations in bold and wh relative clauses on subject position underlined.

*The reporting of a distress incident to a unit which can provide or co-ordinate **assistance**.*

SAR15 displays a D3 score (4.28) similar to Technology/Engineering Academic Prose (4.7) in Biber (1988). The text examines the rescue coordination processes between a distressed submarine and satellite devices (pre, while and post sequence of events).

In D3, pied piping relative clauses constructions are important positive features within the three different forms of relative clauses of this dimension, and in SAR5 wh-relative clauses in object position (0.02), together with pied-ping (0.08) show, despite the low frequency, higher scores than in SAR15, with (0 and 0.04 respectively). SAR15 and SAR05 show the highest means for wh relative clauses in subject position, 0.12 and 0.15, respectively, whereas the means for CMSC19 and CMSC21 are lower, 0.08 and 0.01, respectively. Despite the identical overall frequency mean for phrasal coordination in both corpora, the two SAR texts (SAR05 1.19 and SAR15 1.70) show higher means than the CMSC texts.

IV.1.4. Dimension 4: argumentative orientation

CMSC and SAR corpora share a moderate orientation towards overtly argumentation and a prominent use of modality devices, with Dimension 4 mean scores of -1.27 and

0.95, respectively (Table 1). The score range varies between CMSC18 (-3.61) and CMSC17 (0.52), and between SAR07 (-7.8) and SAR06 (4.64). Figure 6 shows Dimension 4 (D4) mean scores of the corpora analysed.

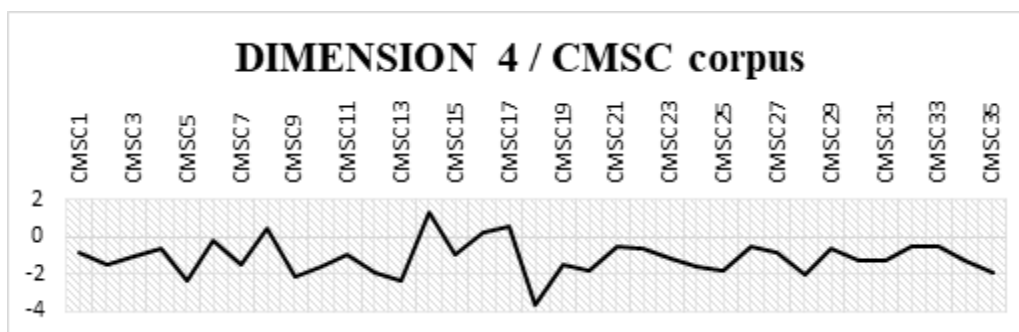


Figure 6. *Dimension 4 scores for the CMSC and SAR corpora*

While professional magazines (CMSC) show a mean score above that of press review texts (-2.8) in Biber (1988), manuals and reports (SAR) show a mean score slightly above phone conversations (0.6). Pairwise comparisons revealed statistically significant differences in D4 scores between the CMSC and SAR corpora $H(2) = 2.863$, $p = .004$.

The defining linguistic features in D4 only display positive weights. They include infinitives, prediction modals (will, would, shall), suasive verbs (agree, ask), conditional subordination, necessity modals (ought to, should, must), split auxiliaries and possibility modals (can, might, may, could) (Biber, 1988). The highest frequency means are observed in conditional subordination in SAR (0.31) versus CMSC texts (0.08), and in the necessity modals in SAR (0.56) versus CMSC texts (0.1). Sample 7 includes CMSC17 and CMSC18 texts. CMSC18 D4 score (-3.61) contributed to the negative overall mean score for the CMSC corpus (-1.27). CMSC18 (2006.1) includes the heavyweight contenders: torpedoes, a text that examines torpedoes as the main hard-kill submarine weapon in the international export markets. It was published in Jane's International Defence Review. CMSC17 (2005.4) The Priz drama describes the Russian mini-sub Priz and her rescue operations with foreign assistance.

Sample 7: CMSC CORPUS

CMSC17 (2005.4) = (0.52)

Infinitives in **bold**.

*They expect the boat s **to grow slightly to improve** the relatively cramped conditions in the existing boats.*

CMSC18 (2006.1) = (-3.61)

Possibility modals and prediction modals underlined.

Its weapon can be installed without integration issues on the Hellenic Navy's new type 214 and upgraded type 209 submarines.

CMSC17 shows higher scores than CMSC18 in most of the relevant linguistic features in D4, which explains its mean score (0.52) and the negative mean score of CMSC18 (-3.61). The linguistic features range from higher values of CMSC17 for infinitives (1.76) vs. CMSC18 (1.40), prediction modals CMSC17 (0.72) vs CMSC18 (0.51) to suasive verbs CMSC17 (0.49) vs CMSC18(0.21) or split auxiliaries in CMSC17 (0.66) vs CMSC18 (0.27). Sample 8 includes SAR6 and SAR7 texts. SAR 6 is entitled *ATP-18_F: Allied manual of Submarine Operations*. This is a 2006 Allied Technical Procedures NATO manual that specifies responsibilities at various levels of command for submarine operations. SAR 7 was introduced in Sample 4.

Sample 8: SAR CORPUS

SAR6 (2006) = (4.64)

Prediction modals underlined.

The submarine will have been instrumental in establishing the maritime superiority in the UWB that will allow the MIO to proceed.

SAR7 (2009) = (-7.8)

Infinitives with to underlined.

The least significant difference test was used for post hoc analysis of the main effect of days at sea to assess day to day changes in alertness.

SAR07 (-7.8) shows general low positive scores. Only infinitives with to (0.72) and suasive verbs (0.35) are higher than those in SAR6, (0.62) and 0.22, respectively. The use of suasive verbs provides intentions to certain actions. These verbs intend to effect a change of some sort (e.g suggest, recommend). Suasive verbs can be followed by a that-clause either with putative should or with a mandative subjunctive. In Sample 8, prediction and possibility modal verbs are frequent. Modality may be defined as the way the meaning of a clause is qualified so as to respect the speaker's judgement of the likelihood of the proposition it expresses being true. Prediction modals (e.g would) are used to discuss hypothetical situations whereas necessity modals (i.e., may) express a plan or intention for certain events.

IV.1.5. Dimension 5: abstract orientation

Both corpora show different Dimension 5 (D5) scores. However, they share an orientation towards abstraction. The score range varies between CMSC20 (1.01) and CMSC31 (3.27), and between SAR07 (-2.2) and SAR14 (3.56). Figures 7 and 8 show D5 mean scores of the corpora analysed.

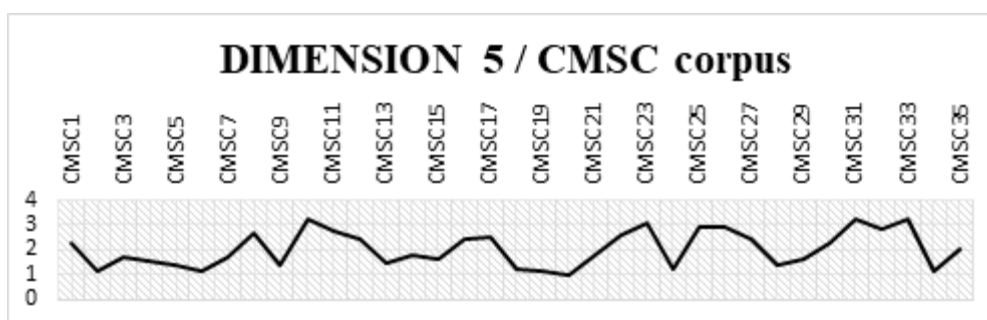


Figure 7. Dimension 5 scores for the CMSC corpus

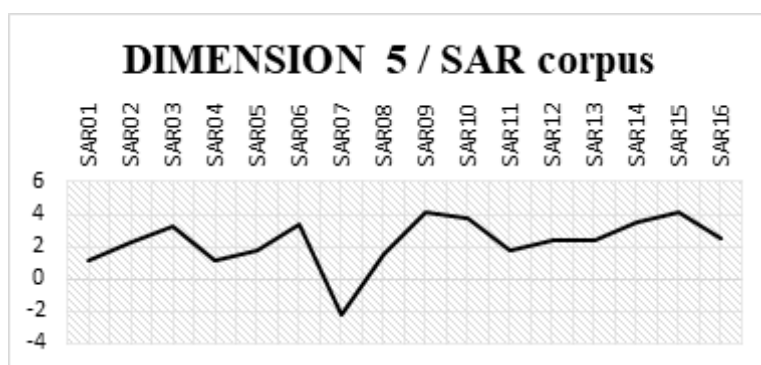


Figure 8. Dimension 5 scores for the SAR corpus

While manuals and reports (SAR) show a mean score of 1.64, professional magazines (CMSC) show a mean score (2.47) above that of press review or hobbies texts (1.2) in Biber (1988). Pairwise comparisons did not show statistically significant differences in D5 scores between the CMSC and SAR corpora $H(2) = 1.076, p = .282$.

The CMSC corpus shows a slightly higher degree of abstractness. Linguistic features that are relevant in D5 include conjuncts (alternatively, altogether, else, etc), agentless passives, adverbial past participial clauses, by-passives and predicative adjectives. The use of agentless passives is the feature that possibly most distinguishes these two corpora: CMSC (1.34) vs. SAR (1.51). The use of the passive voice with agentless passives displays high scores in CMSC. In sample 9 we find CMSC20 and CMSC31 texts. CMSC20 (2006.3) includes *Germany's type 212 A rewards faith in AIP* published in Jane's Navy International. It is a description of a new generation type of German subs which are ready to enter operational service. CMSC31(2010.1&2) is a briefing published in Jane's Defense Weekly journal entitled *Nuclear Deterrent options*.

Sample 9: CMSC CORPUS

CMSC20 (2006.3) = (1.01)

Underlined agentless passive, **conjuncts** in bold.

...which is stored at 180 in tanks under the boat outer skin, but pressure hull, **hence** the submarines increase in size.

CMSC31(2010.1&2) = (3.27)

Underlined agentless passive, **conjuncts** in bold.

Also, while acknowledging that such a capability is not intended to deter terrorist groups.

CMSC31(3.27) shows higher positive weights in all the linguistic features except for adjectives, where is lower in CMSC31(0.50) than in CMSC20 (0.63). Some examples of these higher positive scores can be observed in conjuncts (0.31) vs. CMSC20 (0.20), or agentless passive (1.49) in CMSC31 vs 0.84 in CMSC20. Similarly, past participial clauses have a value of 0.18 in CMSC31 and 0.07 in CMSC20.

Sample 10 includes SAR7 and SAR14 texts. SAR7 can also be found in Sample 8. SAR14 includes *ATP-57.2_EDA_v3: Standards related document. The submarine search and rescue manual*. It is an Allied Tactical Procedures manual published by NATO. In D5, the SAR minimum value is SAR7 (-2.2).

Sample 10: SAR CORPUS

SAR7 (2009) = (-2.2)

Underlined agentless passive, in bold **past participles with deletion**.

*The expected level of performance effectiveness is based upon the detailed analysis of data from participants engaged in the performance of cognitive tasks during several sleep deprivation studies **conducted** by the Army, Air Force and Canadian researches.*

SAR14 (2017) = (3.56)

Underlined agentless passive.

Four pairs of salvage air fittings are located along the top surface of the submarine hull.

IV.1.6. Summary of findings

Table 2 shows the main findings for each of the five Dimensions after the MD analysis of both corpora in this study.

Table 2. Summary of findings

Dimensions	Main finding	Interpretation
Dimension 1 (CMSC – 19.25/ SAR –19.56)	Similar negative mean scores in professional magazines and manuals and reports.	Manuals and reports (SAR) show higher scores in nominalizations, nouns, plain adverbials, agentless passive, and present participles with deletion. It is in nominalizations and nouns where the highest scores are found, which indicates a tendency towards condensed information that contributes to the expression of highly specialised and informational context. Professional magazines (CMSC) show the highest score in type/token ratio, which reflects, according to Biber (1988), a larger diversity of lexical items. Despite the lower mean score in nouns, CMSC texts show the highest score in attributive adjectives.
Dimension 2 (CMSC – 2.9/ SAR – 4.59)	CMSC shows a more narrative orientation.	Manuals and reports (SAR) score suggest a more expository style than CMSC texts, linked to attributive nominal elaboration and immediate time (Biber, 1988). The frequency and distribution of past tenses, third person pronouns, perfect aspect and public verbs in the CMSC is associated with a stronger narrative tendency.
Dimension 3 (CMSC 5.77/SAR 8.12)	SAR texts tend to be more informational	Manuals and reports (SAR) show the highest mean score for nominalization. Professional magazines (CMSC) display lower mean scores (5.7 versus 8.1 in SAR). Time and place adverbials show the lowest scores.
Dimension 4 (CMSC – 1.27/ SAR 0.95)	SAR texts are more persuasive	Manuals and reports (SAR) show the highest scores in conditional subordination, necessity modals and possibility modals. These modals are associated with the ability or necessity for certain events to occur (Biber, 1988). Manual and reports make use of linguistic features that seek to guide the readers.
Dimension 5 (CMSC 2.47/ SAR 1.64)	Scores are not significantly different	Professional magazines (CMSC) and Manuals and reports (SAR) show similar levels of linguistic abstraction. The frequency and distribution of linguistic features in CMSC texts are similar to official documents in Biber (1988). Agentless passives are more frequent in CMSC, though.

V. DISCUSSION

This research examines two corpora (see Section III) that represent two different registers of interest to Navy submariners: professional submarine magazines and salvage and rescue manuals and reports. Although their situational characteristics are diverging in terms of participants and communicative functions, their channel, production circumstances and general topic domain are similar. Given the lack of research in Submarine English, applied linguists and language instructors, however, may be unfamiliar with the linguistic nature of relevant texts when they are appointed to teach such courses.

CMSC was originally compiled and analysed as a response to the lack of research and teaching materials for Navy submariners. In a previous study, Author (2019) used corpus-based analyses of CMSC to shed some light on the complexity of the noun phrase in the corpus and thus inform the selection of vocabulary to be included in the lessons of the subject English for Navy Submariners. Author (2020) went a step forward and used the SAR to devise DDL activities for the teaching of acronyms. Despite these efforts, using a narrow set of linguistics features can only provide limited insight into the complexities of professional communication (Ford et al., 2021). An MD analysis informs both linguistic insights into the functional underpinning of the registers analysed as well as specialized language teaching, offering quantitative data about linguistic variation in any given domain.

Section IV offered an account of the variation that was found in the two corpora analysed across five dimensions of use. The differences were statistically significant in Dimensions 2, 3 and 4, which suggests that the frequency and distribution of some defining linguistic features behaved differently in professional magazines and in manuals and reports. We argue that an understanding of variation continua can only be achieved by attending to what we describe in this paper as converging and diverging patterns of variation in the two corpora analysed across both individual texts and corpora.

Converging patterns of variation show corpora and texts that behave similarly on a given dimension of use. On the contrary, diverging patterns of variation show how the corpora analysed display frequencies and distributions of linguistic features that facilitate distinct functional interpretations on a given Dimension. While research has tended to focus on the differences (Biber, 2019b), and hence on diverging patterns of variation, we note that, in professional and specialized language analysis, the study of converging patterns of variation can impact on the evaluation of the texts that can inform pedagogy a corpus-pedagogy approach.

In the following paragraphs, we will discuss the linguistic characteristics of the two submariner registers relative to Biber's 1988 dimensions of register variation. These characteristics can potentially inform the design of the curriculum and materials (Crosthwaite & Cheung, 2019) for the aforementioned subject. In IV.1, we discuss the Dimensions where differences between the registers were not found. Section IV.2 explores the linguistic differences found across the Dimensions and their potential impact on language teaching.

V.1. Converging orientation in the corpus analysed

Two corpora show a converging orientation when they display no significant differences in the score of a Dimension in the MD analysis. Professional magazines and manuals and reports make use of linguistic features similarly when fulfilling the underlying communicative functions in D1 and D5 (see Tables 1 and 2). However, their participants and specific topics vary (Biber & Conrad, 2009). As Biber (2019b, p.72) put it "the registers themselves have traditionally been treated as discrete categories. Most corpora are organized [into] non-overlapping categories (e.g., fiction, academic prose [...] with individual texts placed into a single category". In other words, both registers and texts could be analysed in a quantitative, continuous situational space "with individual texts being central or peripheral to the situational characteristics of the register" (Biber, 2019b, p. 72).

CMSC and SAR corpora do not display significant differences in the frequency and distribution of features that construct dense information packaging (D1) or the

tendency towards abstraction (D5). While most register analyses have examined differences between corpora, similarities are equally interesting for language instructors and curriculum designers that may need a closer look at the main characteristics of a register. As Hyland (2007, p.162) suggested, grounding curricula and language teaching in the texts that students will have to interact with can only increase the students' "understanding of the ways language is used to create meanings [and] empowers teachers by offering them ways to analyse texts and reflect on the workings of language". In D1, a high similar negative score for nouns in both corpora suggests the presence of very high density of information. Some components of the two corpora yield almost identical mean scores. For example, SAR5 mean noun score is 30.87, while CMSC32 score is 30.89. In D5, the similar mean scores in both corpora are best exemplified in features such as agentless passives, with similar mean scores in both corpora. Agentless passives are usually associated with an abstract style (Conrad and Biber, 2001), so it may be interesting to observe that texts such as CMSC14 (1.47) and SAR2 (1.43) show almost identical mean scores for this feature. Thus, discovering converging patterns of variation across the two registers analysed (see Figure 9) can inform language teachers about how concrete texts behave in the context of a broader corpus.

Manuals and reports (SAR) display higher means of nominalization and nouns. However, some magazine texts similarly show similar high frequencies. By obtaining the mean scores for each of these linguistic features for each of the texts, language instructors will be able to focus on the situational characteristics (Biber & Conrad, 2009; Biber, 2019b) of the different registers, and thus help learners to contextualise the frequency and adequacy of linguistic features across different registers. Consider samples 1 and 2. Despite the similar mean dimension scores of both corpora, the texts in SAR7 (-28.73) and CMSC34 (-24.23) behave in a more similar way in terms of frequency and distribution of nouns and nominalizations than SAR11 (-15.68) and CMSC3 (-15.18). While the differences of the two corpora are not significant, differences across individual texts as shown in samples 1 and 2 can be useful to illustrate specific situational characteristics and understand variation as a continuum rather than an

absolute measure. Some of the target features that could be exploited in the language classroom are found in Figure 9.

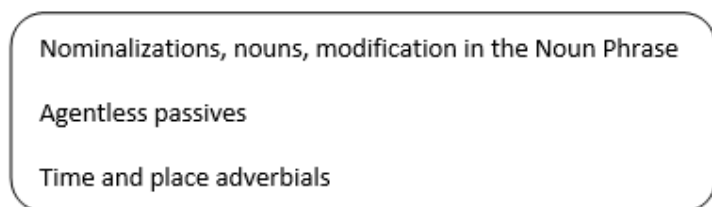


Figure 9. Target linguistic features with similar frequencies and distribution in D1 and D2

V.2. Diverging orientation in the corpus analysed

Two registers show a diverging orientation when they display significant differences in the score of a Dimension in the MD analysis. In this research, significant differences were found on Dimensions 2, 3 and 4 (Tables 1 and 2).

D2 is a good linguistic predictor of register differences (Biber, 2019a) between our two corpora. D2 in the SAR corpus shows the highest negative score (-4.59) versus CMSC (-2.9). Manuals and reports are, expectedly, less narrative than CMSC texts and we can anticipate a lower frequency of features associated with a narrative orientation such as past tenses, third person pronouns or the use of the perfect aspect. More broadly, language instructors could use scores on Dimensions 2, 3 and 4 to select textual evidence of frequency of a given set of linguistic features such as past tenses (D2), time and place adverbials (D3) or modal verbs (D4) that can inform situated uses of linguistic features.

However, it is essential to appreciate that variation across corpora needs to be framed in the context of further variation in individual texts. In other words, it would be wrong to assume that linguistics variation is equally distributed across the corpus components/texts and to approach variation just attending to the general tendency and means in a given corpus. For example, it is counterintuitive to see that in the SAR corpus we find that while SAR1 offers a past tense mean frequency score of 0.34, SAR14 yields a 4.19 mean score, which brings this text closer to the behaviour of past

tenses in professional magazines. Within-corpus variation can thus be useful when illustrating central tendencies, i.e., low frequency of simple past tenses, and uses that diverge from such tendencies. On the other hand, using corpus means can facilitate comparison with other registers and semiotic resources used in different texts. Our MD analysis, for example, confirms that SAR texts tend to behave in a similar way as engineering academic prose, which shows the highest negative score of any register (-4.1) in D2 in (Biber, 1988).

D3 is a good linguistic predictor of register differences in textual elaboration between the two corpora. Textual elaboration is apparent in the SAR corpus. D3 shows explicit textual elaboration through linguistic features with a positive weight such as *wh*-relative clauses in object position, *wh*-relative clauses in subject position, nominalization and phrasal coordination. There are also differences for nominalization in inter-corpus and intra-corpus textual analysis (see samples 5 and 6). D4 is also a good linguistic predictor of register differences.

In D4, manuals and reports show a moderate orientation towards argumentation. In D4 infinitives, prediction modals (*will*, *would*, *shall*), *suasive* verbs (*agree*, *ask*), conditional subordinators, necessity modals and possibility modals (*can*, *may*, *might*, *could*) are relevant linguistic features that can inform corpus-based language teaching and the use of conventional grammatical units of analysis, but again it is essential to bear in mind that features such as necessity modals display diverging frequencies in intra-corpus texts. For example, SAR06 (0.19) and SAR04 (0.71) offer different profiles and different opportunities to examine the occurrence of necessity modals. In sample 8, D4 scores are so diverging that the gap calls for a closer examination of the texts involved. Some of the target features that could be exploited in a corpus-informed curriculum (Hyland, 2007) are found in Figure 10.

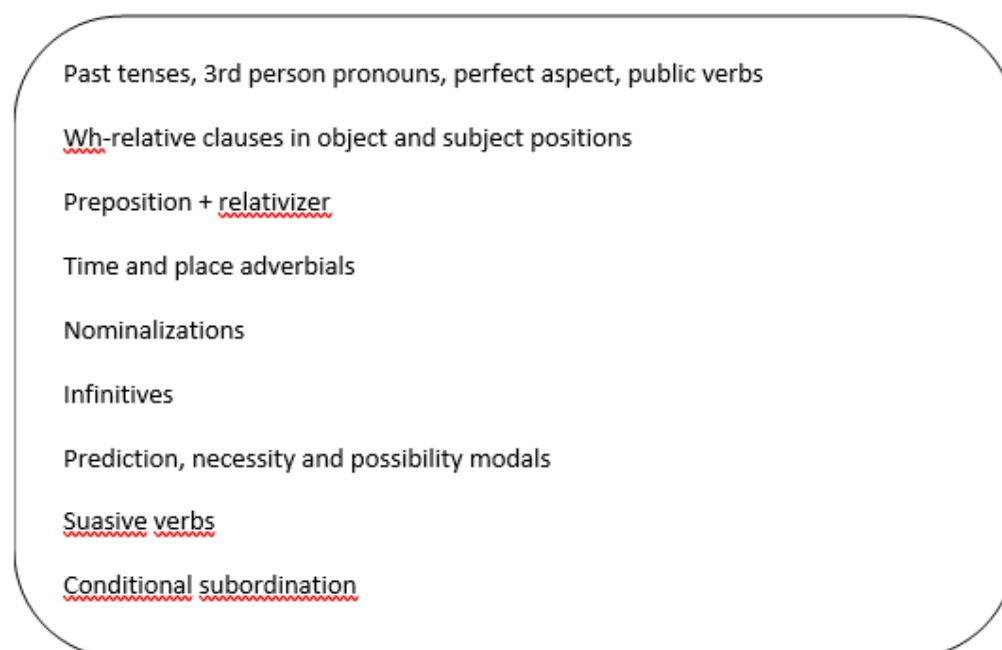


Figure 10. Target linguistic features with diverging frequencies and distribution in D2 -D4

VI. CONCLUSIONS

In this paper, we have operationalized the notion of variation in the context of a MD analysis of two corpora

In this paper, we have operationalized the notion of variation in the context of a MD analysis of two corpora relevant to linguists interested in English for the Military as well as to instructors and students of the subject English for Navy Submariners in the Spanish Navy Submarine Warfare School. We have shown that the analysis of variation can not only inform about the differences between corpora, the default approach according to Biber (2019a), but they can also reveal aspects where corpora show similar patterns of variation. Diverging and converging patterns of variation can therefore provide a fuller linguistic picture of the actual texts used by professionals (Hyland, 2007) and offer instructors the opportunity to use their own data in corpus-based pedagogy (Anthony, 2019).

Similarly, we have provided evidence that intra-corpus variation is equally relevant and needs further attention in LSP pedagogy. Following Biber's (2019b) suggestion, if texts

in a text linguistic register analysis are treated as observations for which rates of occurrence for each linguistic feature are computed, this data can inform about where within-corpus variation can be found, providing valuable information about discursive practices. Understanding how the texts in our data behave on a given Dimension can only provide us with more opportunities to understand how variation works across texts and their situational characteristics.

As Bhatia (2019, p. 47) put it, professional communication needs to be “more efficient in bridging the gap between the academy and the profession, which certainly requires more understanding of and sensitivity to discursive as well as professional practice”. Looking at variation, we note, could inform these much-needed practices, and contribute to bring together corpus-based methods and LSP theory and practice.

Some of the limitations of this study include the use of Biber’s (1988) classic MD analysis framework and the restrictions in place by the Military to access other texts. A new MD analysis of the two corpora may reveal new dimensions of use that are not necessarily identified in this study. Access to classified materials is not, at the time of writing, an option. Further work should examine the use of corpus-based materials that explore the notion of variation and its uptake in a classroom context.

VII. REFERENCES

- Al-Surmi, M. (2012). Authenticity and TV Shows: A Multidimensional Analysis Perspective. *TESOL Quarterly*, 46(4), 671-694.
- Anthony, L. (2019). Tools and strategies for Data-Driven Learning (DDL) in the EAP writing classroom. In *Specialised English* (pp. 179-194). Routledge.
- Bhatia, V. K. (2019). Genre as interdiscursive performance in English for professional. In *Specialised English* (pp. 36-49). Routledge.
- Biber, D. (1988). *Variation across speech and writing*. Cambridge University Press.
- Biber, D. (2008). Corpus-based analyses of discourse: Dimensions of variation in conversation. In K. V. Bhatia, J. Flowerdew and R. Jones (Eds.), *Advances in*

- Discourse Studies* (pp. 100-114). Routledge.
- Biber, D., & Conrad, S. (2009). *Register, genre, and style*. Cambridge University Press.
- Biber, D. (2019a). Multi-dimensional analysis: A historical synopsis. In M. Veirano Pinto and T. Berber-Sardinha (Eds.), *Multi-dimensional analysis: Research methods and current issues* (pp. 11-26). Bloomsbury.
- Biber, D. (2019b). Text-linguistic approaches to register variation. *Register studies*, 1(1), 42-75.
- Biber, D., & Gray, B. (2013). Being Specific about Historical Change: The Influence of Sub-Register. *Journal of English Linguistics*, 41(2).
- Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E., & Quirk, R. (1999). *Longman grammar of spoken and written English* (Vol. 2). Longman.
- Conrad, S. (2001). 4. Corpus linguistic approaches for discourse analysis. *Annual Review of Applied Linguistics*, 22, 75.
- Conrad, S., & Biber, D. (2001). *Variation in English: Multi-Dimensional Studies*. Routledge.
- Ford, J., Paretti, M., Kotys-Schwartz, D., & Howe, S. (2021). New engineers' transfer of communication activities from school to work. *IEEE Transactions on Professional Communication*, 64(2), 105-120.
- Friginal, E. (2009). *The language of outsourced call centers: A corpus-based study of cross-cultural interaction*. John Benjamins.
- Friginal, E. & Roberts, J. (2020). *English in Global Aviation: Context, Research, and Pedagogy*. Bloomsbury Academic.
- Gray, B. (2013). More than discipline: Uncovering multi-dimensional patterns of variation in academic research articles. *Corpora*, 8(2), 153-181.
- Hutchinson, T., & Waters, A. (1987). *English for specific purposes*. Cambridge university press.
- Hyland, K. (2007). English for Specific Purposes. In *International Handbook of English Language Teaching* (pp. 391-402). Springer.
- Nini, A. (2019). The multi-dimensional analysis tagger. In M. Veirano Pinto and T. Berber-Sardinha (Eds.), *Multi-dimensional analysis: Research methods and*

current issues (pp. 67-94). Bloomsbury.

Ren, C., & Lu, X. (2021). A multi-dimensional analysis of the Management's Discussion and Analysis narratives in Chinese and American corporate annual reports. *English for Specific Purposes*, 62.

Sardinha, T. B., & Pinto, M. V. (Eds.). (2019). *Multi-dimensional analysis: Research methods and current issues*. Bloomsbury Publishing.

Received: 02 December 2022

Accepted: 19 December 2022