

Document downloaded from the institutional repository of the University of Alcalá: <http://dspace.uah.es/dspace/>

This is a postprint version of the following published document:

SÁNCHEZ RAMOS, María del Mar; CALZADA PÉREZ, María; "MDA Analysis of translated and non-translated parliamentary discourse". En: Ji, M. y Oakes, M.P. (eds.). *Corpus Exploration of Lexis and Discourse in Translation*. London: Routledge, 2022, p. 26-55 (ISBN: 9781003102694)

<https://www.taylorfrancis.com/chapters/edit/10.4324/9781003102694-2/mda-analysis-translated-non-translated-parliamentarydiscourse-mar%C3%ADa-calzada-p%C3%A9rez-mar%C3%ADadel-mar-s%C3%A1nchez-ramos>
DOI: 10.4324/9781003102694-2

© Routledge (Taylor & Francis Group)



(Article begins on next page)

This work is licensed under a
Creative Commons Attribution-NonCommercial-NoDerivatives
4.0 International License.

MDA analysis of translated and non-translated parliamentary discourse

*María Calzada Pérez, Universitat Jaume I
(Castelló, Spain)*

*María del Mar Sánchez Ramos, University of
Alcalá (Madrid, Spain)*

1 Introduction

Corpus-based translation studies (CTS) are inextricably linked to the name of Mona Baker. It was Baker who, among other things, officially declared the existence of this area of study in a seminal paper published in 1993, in which she foresaw:

a turning point in the history of the discipline. I would like to argue that this turning point will come as a direct consequence of access to large corpora of both original and translated texts, and the development of specific methods and tools for interrogating such corpora in ways which are appropriate to the needs of translation scholars.

(Baker, 1993, p. 235)

Her words anticipated a frantic proliferation of translation-related corpora, which have hitherto undoubtedly enriched translation studies (TS) by providing “new ways of looking at translation” (Kenny, 1998, p. 53). In particular (and among other things), this has increased the “search for patterns that identify translation qua translation” (Laviosa, 2011, p. 18) as part of the identification of what are known as translation universals (Mauranen & Kujamäki, 2004). This has also been furthered in the examination of translator style (Saldanha, 2011).

With almost 30 years since the publication of Baker’s (1993) paper, we can certainly argue that CTS has come of age. We can equally advocate that the time has come to pause for reflection. We are not alone in this view. In a particularly illuminating exercise of self-reflexivity, De Sutter and Lefer (2020) start pondering. They look back into CTS (reasonably mature) history and identify four main problems: a neglect of the notion of similarity and an excessive emphasis on difference; the construction of an underdeveloped theoretical framework that does not follow the stages of

solid, empirical research; the monofactorial nature of research “in which the distribution of a linguistic phenomenon is investigated with reference to one explanatory factor” (De Sutter & Lefer, 2020, p. 5); and the auto-isolation of studies. All of these problems, in their view, result in a reductionist approach that impacts the reliability of the framework, putting into jeopardy its abundant results so far.

De Sutter and Lefer (2020, p. 6) also look ahead and present us with “a new, updated research agenda” in which CTS is to have the following attributes. It is to be multifactorial, embracing the fact that “understanding translation implies understanding its multidimensional structure, and hence multifactorial research designs are essential”. Moreover, it is to be interdisciplinary, and it is to be related to other forms of communication with which it shares processes and products. Finally, it is to be multimethodological; hence, new methods are to be explored for progress to occur.

The present chapter largely shares De Sutter and Lefer’s (2020) declaration of intentions and aims at contributing to the new agenda of these scholars. More specifically, we tentatively intend to enhance CTS’s multifactorial and multimethodological nature through an exploration of Biber’s (1988) interdisciplinary multidimensional analysis (MDA). In building on a multifactorial design, MDA is a clearly underrepresented method in TS. The interdisciplinary nature of its origins (MDA stems from discourse analysis, corpus linguistics, computer studies, and statistics) fights reductionism and exacerbates the complexity of its theoretical basis while adding an empirical slant to its methodology.

In what follows, we have an exploratory go at an MDA study of parliamentary translated and non-translated discourse in English drawing on 2005 language samples from the *European Comparable and Parallel Corpus Archive of European Parliamentary Discourse* (ECPC). Exploring the nature of parliamentary discourse makes sense for its own merits. Parliaments are institutions of the utmost importance to global governance. A wide array of topics (of often great importance for the everyday man and woman) are discussed within its chambers. An ample variety of people (members of parliament, mostly) and styles (informed by ideologies) are confronted and used to pursue similar and dissimilar strategic goals. In the same way, parliaments are a democratic representation of society at large, and parliamentary production is a relatively controlled sample of communication in general. Parliamentary subcorpora may be seen, therefore, to safeguard a form of the representativity that is always so fundamental to Biber’s work. In other words, with its own specifics, of course (parliamentary speech is a genre on its own, after all), one could argue that examining interventions from parliaments is a possible gateway to measuring the state of our societies. Nevertheless, our interest in the ECPC subcorpora here

transcends the parliamentary setting and enters the linguistic arena. European parliamentary houses are impeccably comparable settings for delving into different varieties of translated and non-translated Englishes (such as those from the European Parliament, EP, and the House of Commons, HC). There are other situations of a similar (yet in some respects utterly dissimilar) form (for example, the academic genres analysed by Conrad, 2001). However, the specific cases of the EP (with its translated and non-translated yet possibly “contaminated” Euro-jargon English; for Euro-jargon, see, for instance, Koskinen, 2008, p. 43) and HC (with its relatively independent “pristine” production of “proper” English) offer a plausible scenario for delving into the impact of (translational) contexts on language varieties.

It is important to reinforce that our main motivations through this piece of research are exploratory in nature and humble in their aspirations. However, we see exploration as a necessary inductive (bottom-up) step to establishing solid grounds for upcoming examinations. Our exploration here adopts the following structure. The introduction presented in section 1 is followed in section 2 by a brief contextualization of Biber’s (1988) MDA with some of its main principles and working stages. Section 3 introduces the ECPC sub-corpora, upon which the study is performed together with the main methodological stages used. Section 4 describes the analysis and discusses the results. The chapter ends with concluding remarks.

2 Brief account of Biber’s (1988) multidimensional analysis and its application within TS

2.1 Biber’s (1988) multidimensional analysis

One of the ways in which De Sutter and Lefer’s (2020) new agenda for empirical translation studies may be pursued is by applying (and ultimately adopting) Biber’s (1988) multidimensional analysis (MDA). According to Biber (2014), MDA was born during the 1980s under the influence of two of his mentors at the University of South California: Ed Purcell (who taught him “both statistical analysis as well as advanced computer programming skills”; Biber, 2014, p. 30) and Ed Finegan (his dissertation chair, described by Biber as “central to my development as a corpus linguist, and as a researcher and writer in general”; Biber, 2014, p. 30). Additionally, Biber found inspiration from works by Ervin-Tripp, Firth, Halliday, and Hymes (among others), who noted the importance of correlation for the study of language (Biber, 2019, p. 12). He also drew on research advocating the need for empirical approaches to register variation, such as those presented by Chafe (1982) and Longacre (1976) (cited in Biber, 2019, p. 12). He was especially seduced by Carroll’s (1960; cited in Biber, 2019, p. 12) visionary

study of “vectors of prose style”, employing a statistical analysis of linguistic co-occurrence patterns. That the development of MDA mobilized this ample gamut of interdisciplinary knowledge and skills gives a sense of how very polyhedral and demanding it is.

In brief, MDA is an approach to the study of (monolingual and multilingual) language that ultimately targets “texts, registers, and text types, rather than . . . individual linguistic constructions” (Biber, 1995b, p. 343). Its main ingredients are (1) a corpus-based platform; (2) computational techniques for the automatic identification and disambiguation of linguistic features; (3) multivariate statistics to identify co-occurrence patterns (factors) and relations among texts; and (4) a methodological synthesis of (quantitative) techniques and (qualitative) functional methods, according to which statistical data are interpreted in functional terms. In sum, MDA works under the assumption that:

strong co-occurrence patterns of linguistic features mark underlying functional dimensions. Features do not randomly occur, then it is reasonable to look for an underlying functional influence that encourages their use. In this way, these functions are not posited on an a priori basis; rather they are required to account for the observed co-occurrence patterns among linguistic features.

(Biber, 1988, p. 13)

Most importantly, and according to Biber (1988, p. 20), MDA takes the researcher to a “multi-dimensional space”. Here, analysts do not content themselves with examining data from a one-dimensional perspective by focusing, for example, on a particular language feature within a more or less ample gamut of contexts or, vice versa, by studying multiple features as used in a specific language context/text. Neither is research limited to a two-dimensional prism, whereby two kinds of linguistic items are under scrutiny in the same (more or less varied) number of contexts or according to which a large number of features are dissected within two different settings of language use. Instead, MDA analysts aim at the examination of a large/multiple number of linguistic features in a (more or less wide) multiple range of contexts. Multiplicity is required because, as Biber (1995b, p. 343) argues, “[N]o single linguistic parameter is adequate in itself to capture the range of similarities and differences among spoken and written registers”. Hence, we have the multidimensional label.

Biber’s MDA departs from real data belonging to different (oral and written) genres and moves upwards in four stages: (1) the identification of variables (i.e., language features) to be examined; (2) the extraction of correlations/factors from variables; (3) the functional interpretation of

factors as dimensions; and (4) an overall reflection on relations. The accomplishment of all four stages is what it is called a “Full multidimensional analysis” (Brezina, 2018, p. 161).

In stage 1, analysts must select a list of linguistic features (such as past tense and time adverbials or nominalizations, to name a few for the purposes of illustration) upon which the analysis will be performed. In different implementations of MDA, this list ranges from 40 to 190 items. For example, Biber’s (1988) seminal research works with 67 features (the same ones we use in our study here; for the full list, see Appendix II in Biber, 1988, pp. 221–245). His prior PhD dissertation (Biber, 1984) covers 42 variables. Recent applications of the MDA model have increased this span, with Xiao’s (2009) 141 items and Berber-Sardinha et al. (2014) 190 features.

In stage 2, statistical correlations of the chosen linguistic features are extracted from large pools of corpora (Biber’s, 1988 corpora contained 481 oral and written text samples from 21 different genres). Correlation groupings are known as factors and are identified through the multivariate statistical technique of factor analysis described by Brezina (2018, p. 164) as follows:

a complex mathematical procedure that reduces a large number of linguistic variables to a small number of factors, each combining multiple linguistic variables. This is done by considering correlations between variables . . . ; those that correlate – both positively and negatively – are considered components of the same factor because they have a connection. Positive correlations mean that the variables show the same pattern of occurrence in the data, while negative correlation indicates complementary distribution, that is, if one variable appears with a high frequency the other appears infrequently and vice versa.

In stage 3, MDA pursues connections between factors and language-related situations through the notion of functions. Since all linguistic items are used for a purpose, it is precisely the interpretative analysis of this purpose that serves to make qualitative sense of the statistical data. It is at this stage that correlations/factors become “dimensions”:

Dimensions represent distinct groupings of linguistic features that have been empirically determined to co-occur with significant frequencies in texts. It is important to note that the co-occurring patterns underlying dimensions are identified quantitatively (by statistical factor analysis) and not on any *a priori* basis. Dimensions are subsequently interpreted in terms of the communicative functions shared by the co-occurring features. Interpretative labels are posited for each dimension.

(Biber, 1995b, p. 344)

Biber (1988, chapter 6) identifies six dimensions through which oral and written genres are examined. These six dimensions are among Biber's most important research findings and constitute six different perspectives from which to attempt to produce functional interpretations of the (sub)corpora under study. Due to space constraints, these six dimensions are listed here. Further details about them will be provided as part of our analysis on a need basis. They are:

Dimension 1: "Informational versus Involved Production"

Dimension 2: "Narrative versus Non-narrative Concerns"

Dimension 3: "Explicit versus Situation-Dependent Reference"

Dimension 4: "Overt Expression of Persuasion"

Dimension 5: "Abstract versus Non-Abstract Information"

Dimension 6: "Online Informational Elaboration"

Finally, in stage 4, MDA aims at establishing relations between and among texts (within registers or configurations) and between and among registers (within the linguistic production of speakers/writers or institutions). Relations between and among dimensions may also be scrutinised by MDA, making the approach especially multifaceted.

In sum, Biber's (1988) MDA is a mixed-method research framework that combines quantitative and qualitative components. In addition, it is precisely the latter that, according to Friginal and Hardy (2019, p. 146), constitutes its "real purpose", as they note:

To summarize, although it may seem driven by frequency counts, z-scores, and complex statistics, MD Analysis does not exist in a purely quantitative paradigm. Without qualitative interpretation, there is little meaning to the findings of a linguistic FA.

(Friginal & Hardy, 2019, p. 147)

One qualitative MDA approach involves drawing a comparison to Biber's (1988) original dimensions rather than performing a full MDA. When choosing to do so, researchers skip Stages 1–3 (previously noted) and concentrate on Stage 4 by comparing their results to Biber's material. Relying on Biber's dimensions for qualitative interpretation makes sense since they derive from six factors that "have strong factorial structures, and the features grouped on each factor are functionally coherent and can be readily interpreted on the basis of prior microscopic research" (Biber, 1988, p. 115). This is precisely what we do in section 5. We believe that Biber's well-established dimensions provide a strong interpretative starting point that can inform forthcoming full MDA work, in which we will propose our own factors and dimensions in a more independent manner.

The quantification of real data and their interpretation according to functions, underlying dimensions, and relations make MDA “as much as an art as it is a science . . . it requires technical knowledge of the structure of the language, but it also demands skill, inspiration, and imagination” (Biber, 2014, p. xvi). It is precisely this twofold technical-inspirational/imaginative requirement that makes MDA particularly demanding for researchers, especially in computational and statistical terms.

However, despite its difficulties, MDA has gradually evolved over time with incipient steps represented by Biber (1984) as a direct precursor to the canonical presentation of the MDA model in Biber (1988). Further developments followed suit, especially during the 1990s and beyond. These developments have included, among others, (a) contrastive analyses of different languages (Biber, 1995a); (b) collaborations between Biber himself and PhD students applying MDA to other languages (Besnier, 1988; Biber & Hared, 1992; Kim & Biber, 1994) and (c) diachronic studies describing the evolution of registers (Biber & Finegan, 1992). MDA has transcended Biber’s direct area of influence, and scholars from all over the world have used it to explore a large number of research interests, as Berber-Sardinha and Veirano-Pinto (2014, 2019) testify. That a limited yet reasonable number of research items have been produced within the MDA constellation shows how “powerful” (Brezina, 2018, p. 149) and “reliable” (Nini, 2019, pp. 77–82) MDA is considered to be.

2.2 Multidimensional analysis and translation

Multivariate approaches (such as MDA) have also entered the realm of TS. Admittedly, this has occurred at a slow and hesitant pace, in part due to the difficult technical-inspirational/imaginative requirements mentioned earlier. Nevertheless, with De Sutter and Lefer (2020, p. 1), a handful of other researchers have already approached translation “as an inherently multidimensional linguistic activity and product” (De Sutter & Lefer, 2020, p. 1). In TS, MDA-impregnated projects date at least as far back as Xiao (2010), which focuses on original and translated Chinese within the larger body of translation universals.

Other scholars such as Delaere and De Sutter (2017), Hu et al. (2016), Ji (2017), Kruger (2019), and Kruger and Van Rooy (2016) serve as good examples of how multivariate statistics help go beyond confusing, potentially reductionist, and often unsubstantiated discussions by opening TS’s investigative lens for research. For example, Hu et al. (2016) show that, unlike prior research, translation universals are not to be taken for granted and that empirical translation studies are to proceed from the bottom to the top. Ji (2017) carries out a multifactorial study of universals where different

genres (of translated and non-translated production) are considered at once and where shifts and similarities happen not only locally (among particular features) but also and mainly globally (at the level of genres). Delaere and Sutter (2017), Kruger (2019), and Kruger and Van Rooy (2016) test a larger range of potential causes of the nature of translational texts (such as the bilingual activation mode, the transference of a pre-existing text, cognitive complexity, the SL transfer hypothesis, and above all, pragmatic risk aversion). Kruger and Van Rooy (2016) also bring translation studies to a more open space where they escape autoisolation to establish contact with other forms of “constrained communication” (such as that of the non-native production of English). Finally, De Sutter and Lefer (2020) and especially Kruger (2019) show how very far multivariate studies take us in the empirical examination of previously analysed translational phenomena (such as the implicit/explicit use of “that” with verbs such as “say” or “tell” in translated and non-translated texts). In summary, as a field, TS is problematised to a rather more complex and comprehensive extent when explored with approaches such as MDA.

Studies such as these show that translation is certainly a legitimate (and fruitful) object of study for multidimensional approaches. This is only logical, since MDA’s application gravitates around “the situation”, Biber’s (1988, chapter 2) pivotal notion to justify the existence of language varieties (attending to particular needs or functions) upon which MDA can be performed. Biber (1988) devotes a large part of chapter 2 to the definition of “the situation”, whose components are grounded on Brown and Fraser (1979) and Hymes (1974) (cited in Biber, 1988). Taking these studies as a starting point, Biber (1988) names its main components, which in our view are highly applicable to the case of translation. The first component of all situations is participants’ roles and characteristics. Similar to other situations, there is no doubt that translations have addressors, addressees, and audiences; furthermore, a special type of participant is involved in the translation process, which makes translating unique: the translator or mediator. The second component of such situations pertains to the relations among participants. As in other situations before, during and after translation, participants establish a number of connections (i.e., social, status, or personal relations) with one another. The ample work by Lefevre is seminal in the analysis of, in particular, social and status relations within the translation world. The third collection of situational components includes the setting (where and when communication takes place), topic (what the message is about), and purpose (“outcomes that participants hope for, expect, or intend from the communicative event”; Biber, 1988, p. 32). These three components merge into what Biber (1988) labels “the scene”. There is a translational scene defined as much as any other “constrained” (De Sutter &

Lefer, 2020; Kruger & Van Rooy, 2016) communicative scene. Biber (1988) concludes his examination of situations with a proposal of three pivotal components that are central to the translation process: social evaluation (see sociological approaches in TS); participants' relations to the text – as Munday (2012) shows in his adaptation of Martin and White's (2005) work to translation – and the channel, which through translation give rise at least two subtypes of communication: written translation and oral interpretation. Overall, translations are the result of an urge to communicate specific messages in a particular situation. This is Biber's starting point for deploying MDA procedures and protocols. Consequently, for all the reasons mentioned in this section, we apply MDA to translation in what we consider is a rather innovative way, with the intention of contributing to strengthening the discipline.

3 Methodology

The study described here is an exploratory attempt to perform an MDA on a set of (translated and non-translated) corpora of parliamentary speech in English.

The corpora used for the present analysis belong to the *European Comparable and Parallel Corpus Archive of Parliamentary Discourse* (ECPC). Compiled at Universitat Jaume I (Castellón, Spain), the archive contains (2005–2014) translated and original speeches and writings from (i) the EP in (original and translated) English and Spanish; (ii) the UK House of Commons (HC); and (iii) the Spanish *Congreso de los Diputados* (CD). The specific subcorpora selected for this research contain material in English as follows:

EP_EN_ST_05: Original speeches and written interventions in English from 2005 (870,262 tokens) as published in the *Official Journal of the European Union* (OJEU). This corpus has a standardised type/token ratio (STTR) of 41.20 and a standard deviation (SD) of 58.53. Mean of words per sentence: 22.40.

EP_EN_TT_05: Translated speeches and written interventions into English from 2005 (2,208,677 tokens) as published in the *Official Journal of the European Union* (OJEU). This corpus has a standardised type/token ratio (STTR) of 41.28 and a standard deviation (SD) of 58.69. Mean of words per sentence: 26.27.

HC_05: British MPs' speeches and written interventions from 2005 (7,892,405 tokens) as published in *Hansard*. This corpus has an STTR of 39.20 and an SD of 60.77. Mean of words per sentence: 19.67.

Although we could have used the whole ECPC archive for our study, we opted for a one-year (2005) set of subcorpora for a variety of reasons. First,

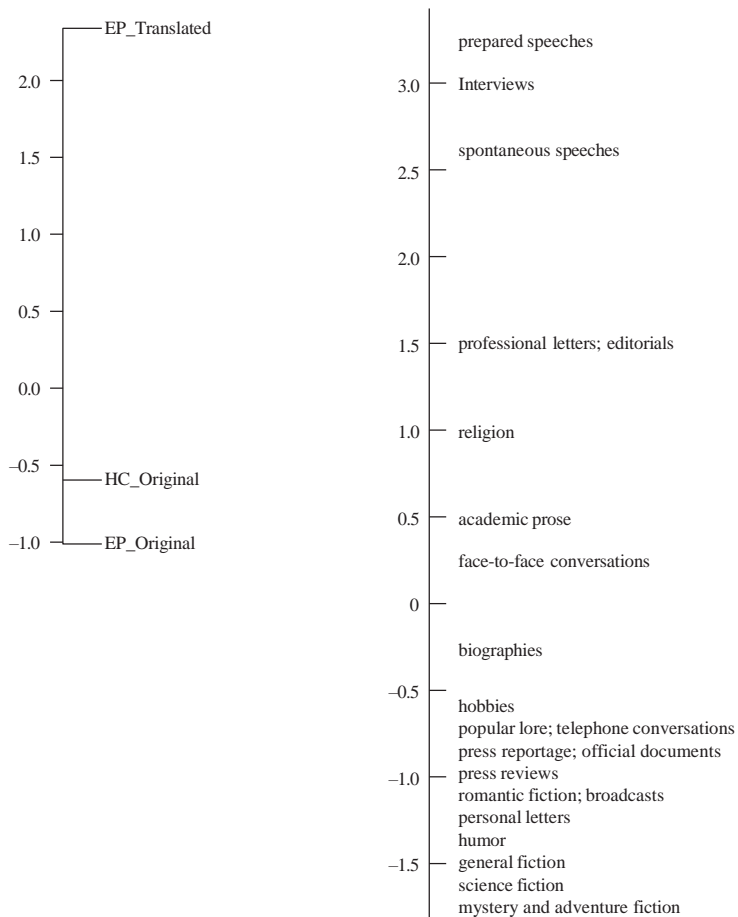
a one-year set of subcorpora seems to be a reasonably large yet manageable amount of linguistic production (totalling 10,971,344 tokens). We see this decision as a compromise between large and small sizes, which helps us with MDA's technical features (too small a corpus may not yield statistically reliable data) while allowing us to control the imaginative requirement as well (too large a corpus will cause researchers to lose sight of details, which often matter for inspirational purposes). Additionally, we conceive of this research as exploratory in nature. We depart with no preconceptions and set off to merely consider the effectiveness of potential synergies between MDA and TS. A set of speeches given in one year seems to be a logical departing point for an exploratory voyage. Finally, our specific focus on 2005 is partly random (2005 is as good or bad as any other year and represents the first full year of speech data from the ECPC Archive) and partly motivated by the fact that it was an eventful year for the EU in which the European Constitution was finally rejected with considerable opposition from the United Kingdom, among others. This frantic activity regarding the Constitution is bound to be reflected by (translated and non-translated) English excerpts of material from the EP and House of Commons. It is then, *a priori*, a good period for exploring similarities and differences.

Hence, in this exploratory study, our departing research questions are kept at a very general (humble, as stated earlier) level and may be phrased in the following manner:

- Can MDA's empirical methodology shed light on the similarities and differences between (translated and non-translated) language in three different types of situations: (a) when original speakers use English for communication at the European Parliament; (b) when translators produce an English version of non-English interventions; and (c) when members of Parliament (MPs) express themselves at the House of Commons? Would it be possible to locate language production in Biberian dimensions and make functional sense of the results?
- Can a TS perspective illuminate and enrich MDA's (largely though not exclusively) monolingual studies?

To answer these questions, drawing mainly on Biber (1988) and reviews by Brezina (2018) and Friginal and Hardy (2019), we focus on the interpretative aspect of Biber's framework and proceed with an analysis involving a comparison to Biber's (1988) MDA results. In future research, we intend to conduct a full quantitative and qualitative study and pursue a full MDA of our corpora.

The present study developed over five stages. In Stage 1, we preprocessed our XML ECPC Archive to revert it to plain TXT format. Each of the



$$F = 47.2; p = 4.031884\text{E-}18, r^2 = 27.5\% \quad F = 8.3; p < .0001; r^2 = 28.5\%$$

Figure 2.1 Example of LSTO results

subcorpora under study includes files containing one-day speeches that have been stripped of all ECPC metadata. With this preprocessing, we could then proceed to Stage 2 and use Nini's MAT (Nini, 2014, 2019). In brief, MAT is a computer programme that (a) produces grammatically annotated versions of TXT (sub)corpora under study using Biber's (1988) 67 linguistic features and (b) generates relative frequencies of these features per hundred words

(among other things). Stage 3 required the use of Brezina's Lancaster Stats Tool Online (LSTO). This is a website where MAT-generated frequencies are inserted to produce graphs of researchers' results alongside Biber's six dimensions, as shown in Figure 2.1.

Note that, if required, the LSTO may also generate numerical and graphical data to perform a full MDA study (which will be useful in forthcoming studies). Stage 4 involved a qualitative analysis of LSTO-produced graphs. In response to the questions posed previously, we aimed to present plausible explanations for ECPC phenomena based on Biber's (1988) dimensions and to complement Biber's framework with ECPC results. In Stage 5, we drew some global conclusions on (original and translated) parliamentary communication from the independently produced examinations of each of Biber's dimensions performed in the previous stage.

4 Analysis and results

As mentioned earlier, the present chapter departs from Biber's MDA in Stage 4 and involves a comparison of his 1988 results to those of EP_EN_ST_05, EP_EN_TT_05, and HC_05. Due to space constraints, we limit the presentation of results to Biber's (1988) dimensions 1, 2, and 6, which are the most prominent for our subcorpora. Hence, the analysis that follows goes through each of these three dimensions. For each dimension, we first present a summary of Biber's interpretation and then examine and exploit its potential to explain our data, which may also contribute to Biber's framework. Before proceeding further, it seems fitting to recall that each dimension reflects a viewpoint from which data may be interpreted.

4.1 Dimension 1

Biber's (1988, p. 107) dimension 1 measures "Informational versus Involved Production". It is a continuum with "high informational density and exact informational content" in the lower end and "affective, interactional, and generalised content" at the top. Thus, texts and registers with low scores on this dimension (e.g., biographies, press reviews, academic prose, press reports, and official documents) are highly informative and lexically precise. They tend to be texts and registers that are less influenced by time/place constraints and that have "careful editing possibilities, enabling precision in lexical choice and an integrated textual structure" (Biber, 1988, p. 107). By contrast, texts and registers with high scores here (e.g., telephone and face-to-face conversations, personal letters, spontaneous speeches, and interviews) occur under circumstances of real-time production and can be seen as "constrained" language (De Sutter & Lefer, 2020; Kruger & Van Rooy,

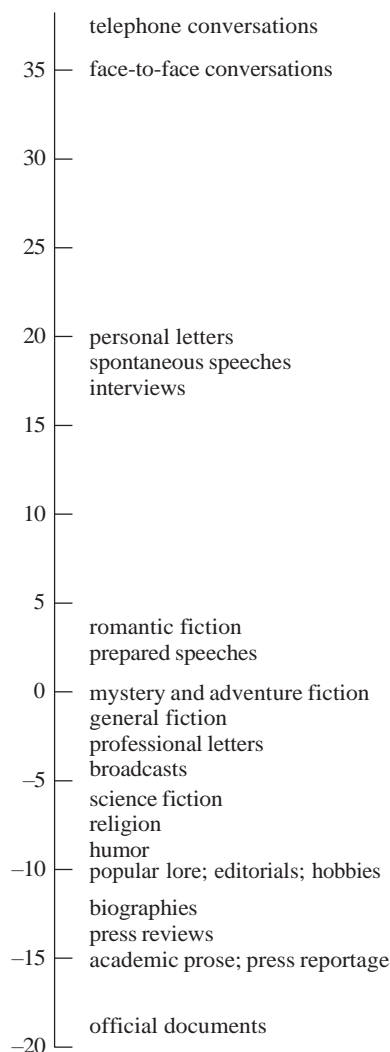
2016) heavily subjected to communicative settings. Editing is not normally possible at this end, and information is presented in a fragmented, more affective manner. In between the top and bottom poles of the dimension are texts and registers with mixed features of informational and involved forms (e.g., romantic fiction, prepared speeches, mystery and adventure fiction, general fiction, professional letters, and broadcasts). Figure 2.2 captures Biber's dimension 1.

Dimension 1 is built from a very long list of positive and negative correlations of features, making it convoluted to interpret. High scores of positive correlation features especially point to the involved (top) end of the dimension. Among them, and for the sake of illustration, we can especially mention private verbs, THAT deletion, contractions, present tense verbs, etc. That is, all of these features appear together in a significant manner and help distribute texts and registers along an involved versus non-involved gradation. High scores of negatively correlated features point to the bottom, informative end of the dimension. Among them, there are nouns, word lengths, prepositions, type/token ratios, etc. Again, all features with negative scores significantly correlate and contribute to distinguishing between more or less informative texts and registers.¹

Dimension 1 is one of the two fundamental dimensions included in Biber's (1988) study (the other being dimension 2, discussed later). Biber (1988) presents solid significance values (F -value = 119.9 and p -value < .0001) and a high Pearson's correlation factor ($r^2 = 84.3\%$). According to Biber (1988, p. 126): "The RxR value indicates the percentage of variation in the dimension scores of texts that can be accounted for by knowing the genre category of the text". In other words, an r^2 of 84.3% means that the dimension 1 parameter accounts for 84.3% of the distribution of the continuum depicted in Figure 2.2. This dimension is then so fundamental that it even qualifies as a candidate "for universal parameters of register variation" (Berber-Sardinha & Veirano-Pinto, 2014, p. xxxiii), which implies that it is supposed to be a determining parameter in distinguishing between and among all (and any) kinds of texts/registers.

However, for our ECPC (original and translated) subcorpora, dimension 1 is largely unimportant (see Figure 2.3).

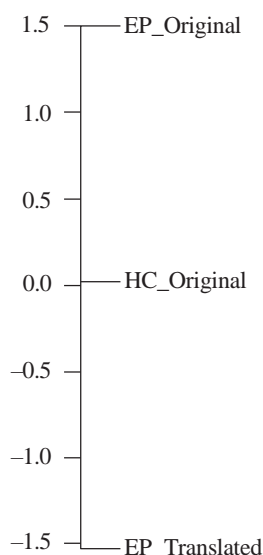
Its r^2 is a mere 2.8%, meaning that it has a dimension explanatory power of just 2.8%. From a different perspective, the figure shows that over 97% of the variation is here explained by factors other than dimension 1. Additionally, these results, though potentially significant in the human science field ($F = 3.6$; $p < 0.05$), are not of the most stringent (e.g., $p < 0.01$ or below). Among others, this may suggest that for this dimension to be more stable, more data may be required. Consequently, regarding dimension 1, the results for the ECPC subcorpora are to be taken with great caution. At any rate, our results may cast doubt over dimension 1 as a strong candidate



$$F = 111.9; p < .0001; r^2 = 84.3\%$$

Figure 2.2 Biber's dimension 1

for universal variation. Even if it has been regarded as a pivotal parameter for distinguishing between and among all texts (it has, after all, some – a 2.8% – significant contributing role in text/register allocation along the dimension), it must be acknowledged that it may have a much lesser impact



$$F = 3.6; p = 0.02781206, r^2 = 2.8\%$$

Figure 2.3 ECPC (original and translated) subcorpora's dimension 1

on results than what Biber's research indicates, depending of course on the kind of study performed.

Taking into consideration all the limitations of dimension 1, we can still say that some of these results are not, however, altogether inexplicable. If dimension 1 is related – among others and as Biber (1988, p. 107) emphasises above – to the possibility of text editing, it is only natural that there is little difference (–1.5 to 1.5 span of variation) between our HC, EP_EN_ST_05, and EP_EN_TT_05 speeches, since all of these are taken from parliamentary proceedings, which are notoriously subjected to proofreading and editing. If editing occurs in all EP and HC cases, then this might partly explain why this parameter proves largely unable to capture similarities and differences among texts and registers. Along the same lines, it is also logical that translated speeches from the EP are more informational than their original EP counterparts since they undergo what may be seen as two rounds of editing: proofing and editing of the original text and then subsequent translation. It is more difficult to explain, however, why House of Commons speeches lie in the middle of our gradation, between translated and original production in EP English.

It is understandable that HC speeches are more exposed to real-time constraints than EP translated interventions due to the very nature of House of Commons debates (where most speakers, for example, are prompted to intervene by the Chair without prior warning and where no translation is required). However, it is not clear why HC speeches are less involved and more informational than original EP speeches. Paradoxically, the latter normally depend on previously scheduled participation by MEPs, who tend to read their interventions in almost all cases. Note again that at the House of Commons, the Chair asks MPs to participate on the spot, and they have very little time to react to the call, adding an element of pressure and improvisation to speech delivery.

Another surprising result is that although the ECPC data have a relatively similar span to that of Biber's prepared speeches (with a score of 2.2), it is remarkably far from his location of spontaneous speeches (with a score of 18.2). The prepared and spontaneous speeches examined by Biber were compiled as part of the 1960s London-Lund corpus. There, prepared speeches include as subgenres: sermons, university lectures, cases made in court, political speeches, and popular lectures. Spontaneous speeches include cases made in court, dinner speeches, radio essays, and speeches made in the House of Commons. Therefore, at first glance, the distance, especially between our HC speeches (at approximately 0.0 in the gradation) and Biber's spontaneous oration (at 18.2 in the gradation), seems baffling. However, we must remember that Biber's oration is directly transcribed by researchers from direct oral interventions while HC speeches are downloaded from already edited *Hansard* proceedings. While on the topic of where ECPC speeches stand along Biber's dimension 1, we should also briefly mention the interesting fact that the EP's original speeches come closer to Biber's (prepared) speeches.

Overall, some of our ECPC results (especially the location of EP translated texts along the continuum) may be (partly) explained by Biber's dimension 1. However, dimension 1 leaves more questions than answers regarding our data, which brings to mind its very low r^2 score. In turn, the ECPC corpora bring further food for thought on Biber's research. Of particular interest and from a translational viewpoint, we could do worse than consider translation as a special form of editing that could tamper with MDA results. Additionally, dimension 1 may well be a universal parameter for distinguishing between and among texts of a very different kind. When texts fulfil an array of similar goals in very different settings (with possibly very different overall aims), as is the case in our research (with EP and HC exchanges), this dimension may lose an enormous proportion of its explicatory capacity. Finally, the a priori most similar types found in ECPC and Biber's corpora and HC and Biber's (spontaneous) speeches are not

necessarily closer in gradation. It is original EP speeches in English that resemble Biber's (prepared) speeches most closely.

4.2 *Dimension 2*

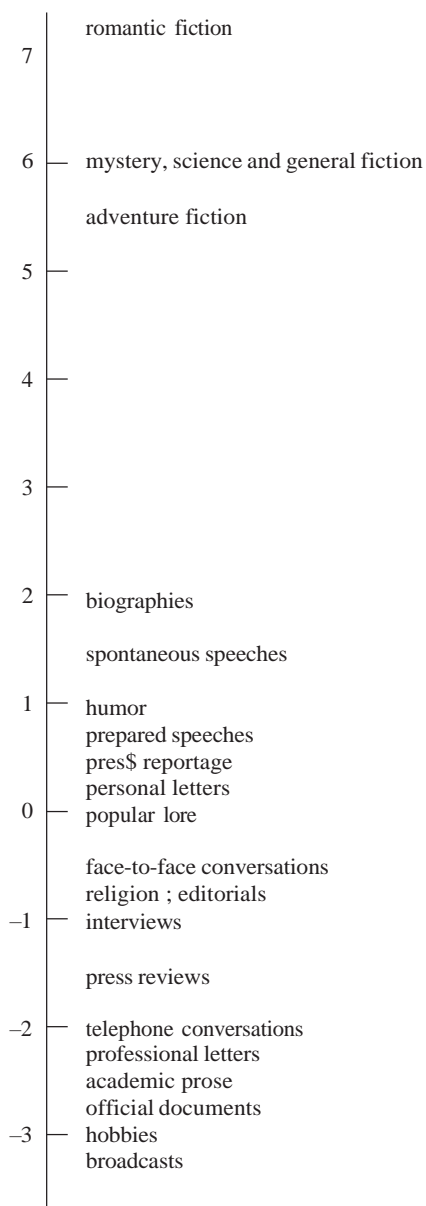
Biber's dimension 2 is "more straightforward than factor 1" (Biber, 1988, p. 108) and measures "Narrative versus Non-narrative Concerns". It thus reflects a continuum that differentiates between narrative discourse from other types of discourse. Texts and registers with high scores on this dimension (e.g., romantic fiction, mystery, science and general fiction, and adventure fiction) normally contain vivid imagery and constantly refer to the past, which frames the story being told. By contrast, texts and registers with low scores here (e.g., press reviews, telephone conversations, professional letters, academic prose, official documents, and broadcasts) tend to focus on the present, dealing "with more immediate matters" (Biber, 1988, p. 109). Between these is a range of genres from biographies and interviews to spontaneous and prepared speeches that display characteristics of both. Figure 2.4 captures Biber's dimension 2.

Dimension 2 is based on a much shorter list of positive and negative correlations of features than dimension 1, which partly explains why it is more straightforward. Positive correlation features include past tense verbs, third person pronouns, perfect aspect verbs, public verbs (introducing reported speech), synthetic negation, and present participial clauses. These features appear together in a significant manner to confer texts a narrative nature. Negative correlated features include present tense verbs, attributive adjectives, past participial deletions, and word length. Again, these features with negative scores significantly correlate to build a non-narrative (often official or professional) discourse.²

Together with dimension 1, dimension 2 is the other candidate as a universal parameter of variation. Biber (1988) presents solid significance values (F -value = 32.3 and p -value < .0001) and a high Pearson's correlation factor (r^2 = 60.8%). Hence, dimension 2 is highly significant and provides 60.8% of the explanation for Figure 2.4, becoming a very strong determining parameter underlying corpora.

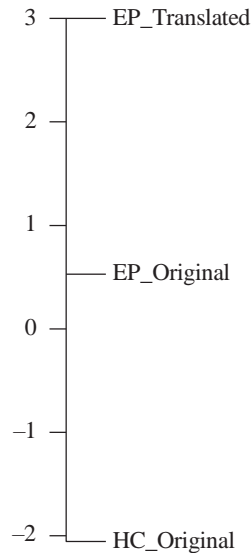
For our ECPC subcorpora, dimension 2 is the most important determining parameter, and 2005's (original and translated) EP and HC parliamentary speeches are displaced on the gradation, as depicted in Figure 2.5.

This arrangement is extremely significant (F -value = 108.7 and p -value < 1.150504E-34) and highly explicative (r^2 = 46.6%) of ECPC corpora variation. Our results thus confirm that dimension 2 is a good candidate for grouping (and separating) texts such as those included in the ECPC subcorpora.



$$F = 32.3; p < .0001; r^2 = 60.8\%$$

Figure 2.4 Biber's dimension 2



$$F = 108.7; p = 1.150504\text{E-}34, r^2 = 46.6\%$$

Figure 2.5 ECPC (original and translated) subcorpora's dimension 2

In our ECPC case, EP speeches in general (regardless of whether they are translated or original) are found to be more narrative than HC interventions. This result is absolutely logical due to the nature and structure of both EP and HC day sessions.

Briefly, and as explained in *The Plenary: a User's Guide* (revision 2019), the Plenary of the European Parliament currently meets four days a month (Monday to Thursday) in Strasbourg. Additional meetings are possible in Brussels. In the past (certainly in 2005, from which texts of our corpora were drawn), the EP normally met between four and six days a month. The Euro Chamber debates and votes on issues carefully included in agendas by the Conference of Presidents, which is composed of the Parliament's President and the leaders of political groups. At the Plenary, most speakers (rapporteurs, MEPs, representatives for the Council and Commission, etc.) are granted a short but fair amount of time (in advance) to advocate their positions in accordance with strict rules of procedures. There is a wide range of potential intervention profiles with their own assigned time slots. Short slots now happen, for example, under the "catch-the-eye" procedure, where MEPs attract the President's attention on a particular topic and, if

granted permission, are allowed to speak for one minute. The catch-the-eye procedure did not exist in 2005, but one-minute interventions were possible at the time under a slightly different procedure.

With this debating structure it comes as no surprise that EP debates excel in narrative concerns. On one hand, the Euro Chamber meets occasionally every month and must consequently look backwards (and forwards) to what happened before the Plenary (and what will happen afterwards). In fact, participants tend to narrate their assessments of past events (with a view to future actions), and to do so they particularly use the features upon which dimension 2 is built. On the other hand, speaking time is often sufficient and carefully preplanned for speakers to develop their own narratives. The following speech (chosen at random from the EP_TT corpus) serves as an example of a typical intervention. Note in particular the number of past tense and perfect aspect verbs used:

Mr President, ladies and gentlemen, this coming 27 January will mark the sixtieth anniversary of the liberation of the concentration camp at Auschwitz. The foundation of the European Coal and Steel Community a few years later **was** a direct consequence of what **had happened** during the Second World War. The fathers and mothers of European cooperation **could** see with their own eyes that ultra-nationalism, Fascism, and the division of our continent by Bolshevism, always **resulted** in oppression, in terror, and in the dignity of the individual being lost. **They could** also see that what **created** peace, dignity and democracy **was** supranational action and integration at a supranational, European level. It is worth adding that this **was** then **achieved** through the Coal and Steel Community gaining control of the weapons-manufacturing industries.

If you trace a continuous line from the ECSC to the European Economic Community, through the next act of integration to the European Community and then to the EU as we know it today, you find a success story as yet unequalled in the world, for it **looked not** only to 1945, and to the end of a war of destruction on this continent; the next stages incorporated states that **had**, in the post-war era, had their own political changes to cope with, namely Greece, Portugal and Spain, and so the overcoming of their fascist dictatorships and the stabilisation of democracy through membership of the European Community **was** an enormous success in the late 1970s and 1980s.

What we saw happening in the 1990s, when the states that **have now acceded took** their present form, was, in principle, a triumph over Communist oppression in one part of Europe. The European Union is a union that **has managed** to take the values described in our Charter

of Fundamental Rights and use them as the basis of democratic integration and to overcome both the Fascism and the Communism of Europe's past.

The events of recent days, in which people **have been** powerless in the face of forces that no amount of technology **has enabled them** to confront, show how absolutely necessary it is for action to be taken at the supranational level, with no national borders standing in the way. In the global village, the European Union is Europe's rational and modern response to the challenges of the twenty-first century. The Constitution we are now debating is the framework provided for it.

Our group will be voting in favour of the Corbett/Méndez de Vigo report – out of profound conviction, and also out of the conviction that the values described in this constitution are civil values. The splendid thing about them is that you can understand them as Christian values if you are a Christian. You can understand them as your own values if you are a Jew or a Moslem or an unbeliever. These values are universal and indivisible, and so they are valid for everyone.

[Martin Schulz on behalf of the PSE group, as translated from German, 11th January, 2005.]

By contrast, the House of Commons meets virtually every day of the year both at the main Chamber and at Westminster Hall. Moreover, a typical form of intervention involves the Chair (Mr/Madam Speaker) allowing MPs to take the floor (for a brief period of time) after they stand up asking to intervene. Short, improvised interventions abound. Everyday meetings seem to justify a discussion of the immediate present or, as *The House of Commons at Work* states, "Many debates relate to day-to-day constituency matters: local hospital services, transport links, planning applications, libraries or an individual's case". Short interventions discourage the narrative presentation of events and tend to encourage quick reactions to critical/laudatory questions about current affairs and especially about the government's performance. An extreme example of this argument is the following intervention, when the speaker (in a very short intervention) discusses not only the immediate present but the very events occurring at the House of Commons:

On a point of order, Mr. Speaker. This is an outrageous abuse. This should be a statement.

[David Maclean (Penrith and The Border) (Con),
10th January 2005]

Why translated speeches are more narrative than original English at the European Parliament begs a closer look that extends beyond the scope of

this exploratory study. However, what is clear at this stage is that original and translated English certainly differ along dimension 2 (by over 2 points in their dimension scores), although not quite as much as with HC speeches (which differ from EP production by over 3 points in their dimension scores). This means that according to our data, the communicative setting has a greater impact on variation than the translational process. Another important finding at this exploratory level is that on Biber's dimension 2, it is (again) original speeches made in English that are closest to Biber's spontaneous (with a dimension score of 1.3) and prepared speeches (with a dimension score of 0.7).

Overall, the similarities and differences observed in our ECPC corpora may be (largely) explained by Biber's dimension 2. Notably, the broad span between the narrative nature of EP and HC speeches identified by the MDA methodology is particularly logical considering the rules of procedures for both chambers. MDA also graphically shows narrative distance between EP original and translated speeches, a finding that requires further research within translation studies. In turn, the ECPC corpora contribute to Biber's research, suggesting greater (original and translated) variation within the parliamentary speech genre in English. Our exploratory research also gives rise to questions left unanswered but that fuel inspiration for further research. Among them are the following. (a) Is there further confirmation of the fact that the narrative nature of speeches is influenced by parliamentary settings to a greater extent than by the translational process? (b) Is there further evidence of the finding that past (1960s) speeches from the HC are more faithfully mirrored by present (2005) speeches from the EP? (c) Could a diachronic study of HC speeches better explore why these differ more from Biber's 1960s parliamentary interventions than original EP speeches? (d) Could the difference between the EP_ST and EP_TT corpora suggest a different conception of narration in parliamentary settings or of the role of speakers in parliaments rather than a change in the narrative content of source and target texts?

4.3 Dimension 6

Biber's (1988, p. 107) dimension 6 measures "Online Informational Elaboration". As seen, dimension 6 is related to dimension 1, and we might even conclude that dimension 6 narrows the prism through which linguistic phenomena are seen in dimension 1, weighing down information of especially those texts produced under constrained circumstances. Thus, texts and registers with low scores on this dimension (e.g., personal letter, humour, general fiction, science fiction, mystery and adventure fiction) are viewed as heavily informative and minimally subject to contextual, spontaneous pressures. By contrast, texts and registers with high scores (notably both

prepared and spontaneous speeches) are here seen as highly informative but also highly constrained by real-time production circumstances (see De Sutter & Lefer, 2020; Kruger & Van Rooy, 2016). Figure 2.6 captures Biber's dimension 6.

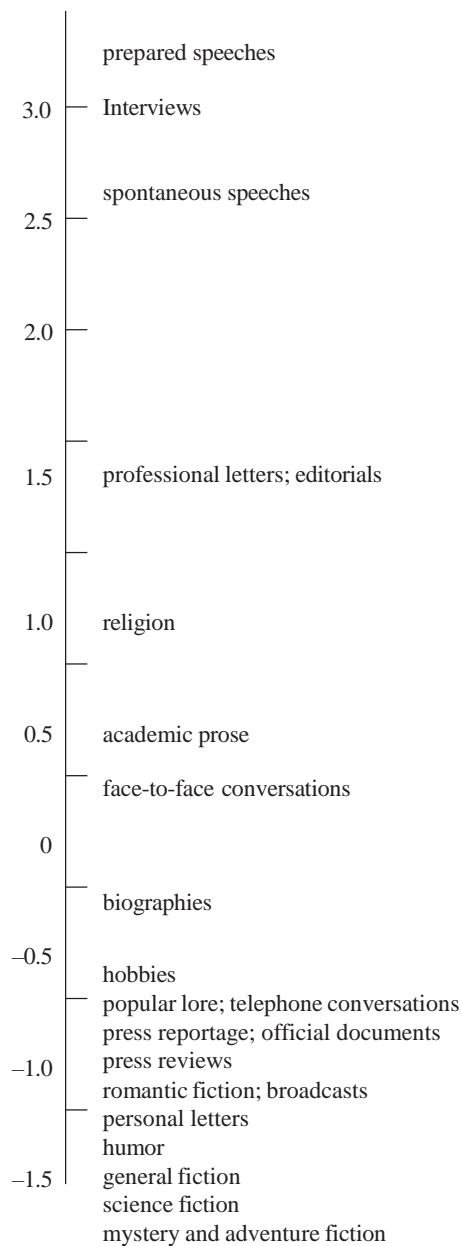
Dimension 6 is based on a very concise list of mainly positive correlation features especially associated with the use of different types of THAT: that clauses as verb complements; that relative clauses; that clauses as adjective complements; and demonstrative pronouns (including that).³

In Biber (1988), dimension 6 has strong significance values (F -value = 8.3 and p -value < .0001) and a reasonably high Pearson's correlation factor ($r^2 = 28.5\%$), providing almost a third of the explanation for the arrangement of Figure 2.6. However, Biber (1988, p. 114) find its interpretation to be difficult at times and ends up acknowledging that it requires "further investigation".

For our ECPC corpora, dimension 6 is the second most important determining parameter (after dimension 2). From its vantage point, 2005 (original and translated) EP and HC parliamentary speeches are placed on the gradation depicted in Figure 2.7.

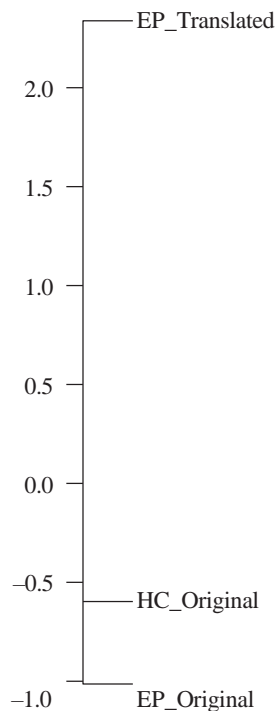
The arrangement of our ECPC corpora along dimension 6 is extremely significant (F -value = 47.2; p -value < 4.031884E-18) and echoes Biber's correlation results ($r^2 = 27.5\%$). These results indicate that almost one-third of the explanations supporting Figure 2.7 are provided by this parameter and that this result is extremely reliable. It might be argued that for TS, dimension 6 has special value due to its connection to the use of (the first type of) THAT, a linguistic feature that has been widely studied in prior research. On the one hand, studies such as Olohan and Baker's (2000) seminal work propose the identification of a much more frequent use of this type of THAT in translations than in original texts. On the other hand, studies such as Delaere and De Sutter (2017) and De Sutter and Lefer (2020) call for more complex (multifactorial) approaches to this object of study.

In our case, the uses and correlations of different types of THAT in translated speeches (from the EP) are clearly and significantly different from those found in original speeches (from the EP and HC). Translations excel in the presence of THAT, while originals rank significantly lower in this respect. Furthermore, in rereading results from dimension 1, which measures, among others, the deletion of THAT in verb complement clauses (e.g., with verbs such as "say" or "tell"), we find that originals from the EP and HC rank higher (and behave more similar to each other) than uses from EP translations. Finally, Biber (1988, p. 243) characterises THAT deletion as "dispreferred in edited writing", relating it to what are seen as unorthodox linguistic constructions.



$$F = 8.3; p < .0001; r^2 = 28.5\%$$

Figure 2.6 Biber's dimension 6



$$F = 47.2; p = 4.031884\text{E-}18, r^2 = 27.5\%$$

Figure 2.7 ECPC (original and translated) subcorpora's dimension 6

Overall, the similarities and differences found in our ECPC corpora may be (largely) explained by Biber's dimension 6. Biber's MDA dimension 6 (and dimension 1) identifies a highly prominent similarity between EP and HC originals in the uses and correlations of different types of THAT regardless of the parliament where speeches are delivered (confirming Olohan & Baker's, 2000 results). Nevertheless, the inclusion of a variety of THAT types in this dimension (and not just in complement to verbs alone) suggests, similar to Delaere and De Sutter (2017) and De Sutter and Lefer (2020), that the issue is more complex than merely the use/deletion of THAT clauses as verb complements and implies that a multifactorial approach, if pursued further, may provide further illustrative explanations for the phenomena. In turn, the ECPC corpora complement Biber's fuzzy and unstable interpretation of dimension 6, which may now be seen under the light of

“dispreferred” (Biber, 1988, p. 243) structures in more or less constrained forms of communication. Moreover, as Biber (1988) argues, all dimensions are inextricably connected. The ECPC corpora provide evidence suggesting that dimensions 1 and 6 are closely linked, and results of the former may be elucidated by phenomena detected in the latter, providing new investigative avenues for global analysis, as is the case here. In reviewing dimension 6, the allocation of subcorpora on dimension 1 (which is very closely related to dimension 6) seems more comprehensible than what we initially gathered.

5 Conclusion

With an exploratory and humble spirit, this chapter has compared MDA results for the ECPC Archive (of 2005 original and translated speeches from the EP and HC) to Biber’s (1988) seminal research. Due to space constraints, the chapter only comments on three (out of the six) dimensions Biber identifies in 1988 (i.e., dimensions 1, 2 and 6), which prove more illustrative for our exploratory purposes. We set off with the intention to test whether Biber’s methodology offers plausible explanations for (original and translated) parliamentary speeches and whether original and translated material could produce valuable insights for Biber’s framework. This is performed with a view to contribute to De Sutter and Lefer’s (2020) new agenda for empirical translation studies.

Even at this exploratory stage, this chapter suggests that MDA may be seen as a valuable tool for new TS agendas. On each dimension, similarities and differences between the subcorpora are identified. Functional interpretations are then made with what we believe is a high degree of exploratory success. When we gather results from all dimensions, we may build a global account of the texts under study (which are to be pursued further in future research).

In our study, EP_TT texts were found to be particularly informational and narrative and to excel in the orthodox (linguistically preferred) use of THAT (of various types). EP_ST texts were found to be particularly involved and to opt for the unorthodox/linguistically dispreferred use of THAT (of various types). Finally, HC texts appeared as particularly non-narrative and occupied a middle ground (between EP_TT and EP_ST) as per levels of involvement and a preference for orthodoxy in the use of THAT (of various types). As is shown, the translation process could be said to play an important part in the allocation of subcorpora along the dimensions (see dimensions 1 and 6). By contrast, sometimes, it is the context of communicative exchange (the EP or HC in our case) that seems to have a greater impact (as in dimension 2). Plausible reasons for these results are discussed throughout section 4.

In conclusion, from this exploration, synergy between MDA and TS has indisputable advantages, some of which are as follows:

- 1 Graphically (and statistically) locating subcorpora along each of Biber's dimensions
- 2 Characterising subcorpora globally with information of all dimensions
- 3 Identifying objective data that are then to be interpreted with subjectivity based on the overall context of communication

Nevertheless, MDA is certainly not the be-all and end-all of a new agenda for TS for a variety of reasons, among which the following come to mind:

- 1 (Some) MDA statistical information is not to be regarded as infallible. An r^2 of 2.8% (as in our dimension 1), for example, shows that the results are influenced by a wide range of other possible causes left unstudied by our MDA. However, we equally believe that an r^2 of 2.8% does not mean that the results are to be fully discarded since, after all, they do have a statistically significant impact on the phenomena under study.
- 2 MDA quantitative data are insufficient to gauge linguistic phenomena on their own; a qualitative (inspiration/imaginative) examination of contexts will be as important in researching in the direction to which statistics point.

Acknowledgements

The present research is funded by the Ministry of Science and Innovation in Spain (PID2019–108866RB–100).

We would like to express our gratitude to Vaclav Brezina for introducing us to the most technical aspects of multidimensional analysis.

Notes

- 1 See Biber (1988, pp. 102–103) for a complete summary of factorial structure.
- 2 See Biber (1988, pp. 102–103) for a complete summary of factorial structure.
- 3 See Biber (1988, pp. 102–103) for a complete summary of factorial structure.

References

- Baker, M. (1993). Corpus linguistics and translation studies: Implications and applications. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), *Text and technology* (pp. 233–250). Amsterdam: John Benjamins Publishing Company.
- Berber-Sardinha, T., Kauffman, C., & Mayer-Acunzo, C. (2014). Dimensions of register variation in Brazilian Portuguese. In T. Berber-Sardinha & M. Veirano-Pinto

- (Eds.), *Multi-dimensional analysis, 25 years on: A tribute to Douglas Biber* (pp. 35–79). Amsterdam: John Benjamins Publishing Company.
- Berber-Sardinha, T., & Veirano-Pinto, M. (Eds.). (2014). *Multi-dimensional analysis, 25 years on: A tribute to Douglas Biber*. Amsterdam: John Benjamins Publishing Company.
- Berber-Sardinha, T., & Veirano-Pinto, M. (Eds.). (2019). *Multi-dimensional analysis: Research methods and current issues*. London: Bloomsbury Academic.
- Besnier, N. (1988). The linguistic relationships of spoken and written nukulaelae registers. *Language*, 64(4), 707–736. <https://doi.org/10.2307/414565>
- Biber, D. (1984). *A model of textual relations within the written and spoken modes* (Unpublished doctoral dissertation). Los Angeles: University of Southern California.
- Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Biber, D. (1995a). *Dimensions of register variation: A cross-linguistic comparison*. Cambridge: Cambridge University Press.
- Biber, D. (1995b). On the role of computational, statistical, and interpretive techniques in multi-dimensional analyses of register variation: A reply to Watson. *Text: Interdisciplinary Journal for the Study of Discourse*, 15(3), 341–370. <https://doi.org/10.1515/text.1.1995.15.3.341>
- Biber, D. (2014). Multidimensional analysis: A personal history. In T. Berber Sardinha & M. Veirano Pinto (Eds.), *Multi-dimensional analysis, 25 years on: A tribute to Douglas Biber* (pp. xxix–xxxviii). Amsterdam: John Benjamins Publishing Company.
- Biber, D. (2019). Multi-dimensional analysis: A historical synopsis. In T. Berber Sardinha & M. Veirano Pinto (Eds.), *Multi-dimensional analysis: Research methods and current issues* (pp. 11–26). London: Bloomsbury Academic.
- Biber, D., & Finegan, E. (1992). The linguistic evolution of five written and speech-based English genres from the 17th to the 20th centuries. In M. Rissanen, O. Ihalainen, & T. Nevalainen (Eds.), *History of Englishes: New methods and interpretations in historical linguistics* (pp. 688–704). Berlin: Mouton de Gruyter.
- Biber, D., & Hared, M. (1992). Dimensions of register variation in Somali. *Language Variation and Change*, 4(1), 41–75. <https://doi.org/10.1017/S095439450000065X>
- Brezina, V. (2018). *Statistics in corpus linguistics: A practical guide*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781316410899>
- Conrad, S. (2001). Variation among disciplinary texts: A comparison of textbooks and journal articles in biology and history. In S. Conrad & D. Biber (Eds.), *Variation in English: Multi-dimensional studies* (pp. 94–107). Harlow: Pearson.
- Delaere, I., & De Sutter, G. (2017). Variability of English loanword use in Belgian Dutch translations: Measuring the effect of source language and register. In G. De Sutter, M. A. Lefer, & I. Delaere (Eds.), *Empirical translation studies: New methodological and theoretical traditions* (pp. 81–112). Berlin: Mouton De Gruyter.
- De Sutter, G., & Lefer, M. A. (2020). On the Need for a new research agenda for corpus-based translation studies: A multi-methodological, multifactorial and interdisciplinary approach. *Perspectives*, 28(1), 1–23. <https://doi.org/10.1080/0907676X.2019.1611891>

- European Parliament. (2019). *The plenary: A user's guide*. www.europarl.europa.eu/sed/doc/ext/manual/Plenary_guide_en.pdf
- Friginal, E., & Hardy, J. A. (2019). From factors to dimensions: Interpreting linguistic co-occurrence patterns. In T. Berber Sardinha & M. Veirano Pinto (Eds.), *Multi-dimensional analysis: Research methods and current issues* (pp. 145–164). London: Bloomsbury Academic.
- House of Commons. (2018). *House of commons at work*. www.parliament.uk/documents/commons-information-office/Publications-2015/House-of-Commons-at-work-booklet.pdf
- Hu, X., Xiao, R., & Hardie, A. (2016). How do English translations differ from non-translated English writings? A multi-feature statistical model for linguistic variation analysis. *Corpus Linguistics and Linguistic Theory*, 15(2), 347–382.
- Ji, M. (2017). A multidimensional analysis of the translational Chinese genre system. In M. Ji, L. Hereide, D. Li, & M. Oake (Eds.), *Corpus methodologies explained: An empirical approach to translation studies* (pp. 53–102). London: Routledge.
- Kenny, D. (1998). Corpora. In M. Baker (Ed.), *Routledge encyclopedia of translation studies* (pp. 50–53). London: Routledge.
- Kim, Y., & Biber, D. (1994). A corpus-based analysis of register variation in Korean. In D. Biber & E. Finegan (Eds.), *Sociolinguistic perspectives on register* (pp. 157–181). Oxford: Oxford University Press.
- Koskinen, K. (2008). *Translating institutions: An ethnographic study of EU translation*. Manchester: St. Jerome.
- Kruger, H. (2019). That again: A multivariate analysis of the factors conditioning syntactic explicitness in translated English. *Across Languages and Cultures*, 20(1), 1–33.
- Kruger, H., & Van Rooy, B. (2016). Constrained language: A multidimensional analysis of translated English and a non-native indigenised variety of English. *English World-Wide*, 37(1), 26–57.
- Laviosa, S. (2011). Corpus-based translation studies: Where does it come from? Where is it going? In Kruger, K. Wallmach, & J. Munday (Eds.), *Corpus-based translation studies: Research and applications* (pp. 13–32). London: Continuum.
- Martin, J. R., & White, P. R. R. (2005). *The Language of Evaluation. Appraisal in English*. Basingstoke: Palgrave MacMillan.
- Mauranen, A., & Kujamäki, P. (2004). *Translation universals do they exist?* Amsterdam: John Benjamins Publishing Company.
- Munday, J. (2012). *Evaluation in translation: Critical points of translator decision-making*. London: Routledge.
- Nini, A. (2014). *Multidimensional analysis tagger 1.2: Manual*. <http://sites.google.com/site/multidimensionaltagger>
- Nini, A. (2019). The multi-dimensional analysis tagger. In T. Berber Sardinha & M. Veirano Pinto (Eds.), *Multi-dimensional analysis: Research methods and current issues* (pp. 67–94). London: Bloomsbury Academic.
- Olohan, M., & Baker, M. (2000). Reporting that in translated English: Evidence for subconscious processes of explicitation? *Across Languages and Cultures*, 1(2), 141–158.

- Saldanha, G. (2011). Translator style: Methodological considerations. *The Translator*, 17(1), 25–50. <https://doi.org/10.1080/13556509.2011.10799478>
- Xiao, R. (2009). Multidimensional analysis and the study of world Englishes. *World Englishes*, 28(4), 421–450. <https://doi.org/10.1111/j.1467-971X.2009.01606.x>
- Xiao, R. (2010). How different is translated Chinese from native Chinese? A corpus-based study of translation universals. *International Journal of Corpus Linguistics*, 15(1), 5–35.