What is kept and what is lost without translation? A Corpus-Assisted Discourse Study of the European Parliament's original and translated English

María Calzada Pérez

University Jaume I, Castellón de la Plana, Spain

calzada@uji.es

Abstract

In July 2011, the European Parliament (EP) stopped providing a written translation of its proceedings. Some years later, it seems apposite to look back and ask: What is kept and what is lost without the EP translating? To answer this question, the present paper adopts the first (Modern Diachronic) Corpus-Assisted Discourse Analysis Study (MD-CADS) carried out within Translation Studies by drawing on the Discourse-Historical Approach (DHA) and Corpus Linguistics (CL) tools. Hence, along DHA lines, the paper proceeds from texture through strategies to content by focusing on CL key keywords and detailed consistency. It performs analysis upon the European Comparable and Parallel Corpus (ECPC) archive, compiled at the Universitat Jaume I (Spain). This study shows that MD-CADS is a potential source of data for triangulation with other more qualitative approaches.

Keywords: translation, MD-CADS, DHA, CL, European Parliament

Introduction

Europe and translation are inextricably related. If, for Bauman (2004, p. 89), 'Europe has been and remains a homeland of perpetual translation', the famous sentence 'translation is the language of Europe', traditionally attributed to Umberto Eco, is indeed profusely affirmed within European Union (EU) institutions.[1] Much of what is discussed within the framework of the EU is done via translation or interpreting, and many of the messages that come from Euro settings are, to a large extent, translated messages. This was certainly the case for the European Parliament's (EP) plenary speeches until July 2011, when the decision was taken to stop producing written translations of its proceedings. A (much cheaper) original-language-only 'rainbow version' is now published instead, though linked to videos containing the original audio of speakers and interpreters. But the change of policy regarding the written version allows an interesting study of variability in translated and non-translated discourse before and after 2011.

By analysing what happened when EP proceedings were translated (vis-à-vis comparable original rendering at the time) we can start a discussion about the potential consequences of the interruption of translation. I address this topic from a (Modern Diachronic) Corpus-Assisted Discourse Study perspective (MD-CADS). In doing so, I

---

[1] See for example http://www.eutrio.be/language-europe-translation,
http://cdt.europa.eu/CDT%20Publication%20Book/The%20language%20of%20Europe%20-
%20Translating%20for%20the%20EU/THE%20LANGUAGE%20OF%20EUROPE.pdf

adopt some of the premises of the Discourse-Historical Approach (DHA) while advocating the robustness of Critical Linguistics (CL) tools.

1. Framework of Analysis: MD-CADS, DHA and CL

MD-CADS

For some time now, Corpus-Assisted Discourse Studies (otherwise known as CADS, see Partington 2004) has been putting corpus linguistic tools and methodologies at the service of discourse analysis, with what have been seen to be 'impressive results' (Garzone & Santulli, 2004, p. 353). In particular, Modern Diachronic Corpus-Assisted Discourse Studies (MD-CADS) generates a type of CADS work dedicated to illustrating

> […] how two or more 'sister corpora', that is, corpora constructed to be as similar in content, composition and structure as possible, might enable an analyst to track how language patterns and meanings, as well as the discourse practices these might reveal, can change over comparatively brief periods of modern times […] (Partington, Duguid, & Taylor, 2013, p. 265).

(MD-)CADS analysts (e.g. Baker, 2010; Baker et al., 2008; Baker & McEnery, 2005; Garzone & Santulli, 2004; Hardt-Mautner, 1995; Koller & Mautner, 2004; Mautner, 2009a, 2009b; Mehan, 1997; O'Halloran & Coffin, 2004; Orpin, 2005; Partington et al., 2013; Stubbs, 1996) are adamant in showing that descriptive, (semi-)automatic, quantitative studies are not necessarily at odds with interpretative, manual and qualitative research.

Much of (MD-)CADS-inspired synergic research either focuses on original texts or does not take translation into account. Furthermore, to the best of the present author's knowledge, only a handful of studies apply corpus-based methodologies to the examination of parliamentary communication (e.g., Baker, 2006, 2010; Bayley, 2004; Carpuat, 2014). There is no modern diachronic study of the communicative dynamics of the EP. However, much of Europe's most important decision-making occurs in the Euro-Chamber through multilingual exchanges, and its agenda influences ideologies, behaviours and the language of national houses. The present article therefore aims to contribute to filling this twofold (translational and diachronic) gap in study.

In order to perform (MD-)CADS analyses, '[t]he major problem', as Orpin (2005, p. 38) admits, 'is deciding where to start'. More specifically, researchers must adopt a specific line of discourse analysis and choose concrete Corpus Linguistics (CL) tools to perform their work. These are the first two issues discussed below.

DHA

Among the many Discourse Analysis perspectives that may be adopted within (MD-)CADS, the present study opts for Wodak and Reisigl's Discourse-Historical Approach (DHA). The reason for my choice is that DHA has already been put to the test in the

examination of European phenomena (e.g. Barenreuter, 2005; Krzyżanowski, 2005; Wodak, 2011), and it seems the sensible option for the purpose of my study. Arguably, DHA is especially suited to 'go beyond studies critical of the present and aim at comparative analysis both in diachronic and intercultural terms' (Wodak, 2011, p. 7). These two (diachronic and intercultural) axes are indeed essential for an (MD-)CADS comparison of original and translated EP speeches, such as proposed here.

DHA advocates a top-down analysis that starts with an exhaustive ethnographic examination of the historic and generic contexts in which the texts under discussion are produced. Then researchers turn to the actual texts and, by analysing them meticulously, sentence by sentence, they move from means and forms of realization through strategies to content, which they see as closely associated with the context already studied (see Wodak, De Cillia, Reisigl, & Liebhart, 1999, pp. 36–42). Contents, strategies and means are three analytical dimensions that are 'closely interwoven' (Wodak et al., 1999, p. 30) and particularly relevant to my work here.

The content dimension is straightforward and refers to the thematic areas of the objects of study. Means and forms of realization are also easy to comprehend since they refer to the different linguistic features (or textural traits) that make up texts. DHA's strategies, however, require further explanation. In a Bordieusian fashion, '[s]trategic action is oriented towards a goal but not necessarily planned to the last detail or strictly instrumentalist; strategies can also be applied automatically' (Wodak et al., 1999, p. 32). DHA identifies strategies of various kinds and at various (more or less abstract) levels of application. For the purposes of this study, I highlight the operationality of the following:

1. Nomination: 'discursive construction of social actors, objects/phenomena/events and processes/actions' (Wodak & Meyer, 2009, p. 94). This strategy seems to take place within the area covered by Halliday's (1985) ideational meaning and, more specifically, in relation to participants and processes.

2. Predication: 'discursive qualification of social actors, objects, phenomena, events/processes and actions (more or less positively or negatively)' (Wodak & Meyer, 2009, p. 94). Adjectives and other types of modifiers (such as appositions, relative clauses, prepositional phrases, and so on) are the means to convey this strategy.

3. Argumentation: 'justification and questioning claims of truth and normative rightness' (Wodak & Meyer, 2009, p. 94). Fallacies and linguistic *topoi* are some of the most usual notions behind argumentation (for a detailed introduction of argumentation see, for example, Eemeren & Houtlosser, 2006).

4. Perspectivation: 'positioning speaker's or writer's point of view and expressing involvement or distance' (Wodak & Meyer, 2009, p. 94). Pronouns and direct and indirect speech are among those forms that point toward text perspective.

5. Intensification and mitigation: 'modifying (intensifying or mitigating) the illocutionary force and the epistemic or deontic status of utterances' (Wodak & Meyer, 2009, p. 94). Modals and verbs of saying, thinking, and feeling are tools of this strategy.

With DHA, then, we gain three linguistic dimensions (contents, strategies, and means) of analysis for an (MD-)CADS methodology. However, this article changes the direction and emphasis of study, opting for an inductive methodology that takes us from the bottom (the focus of my analysis) to the top. I also propose to deal with the means of realization in a more automatized and quantitative manner. At the same time, and due to space constraints, I leave the content-context connection for further research. My decision is here presented as compatible and complementary to that of DHA's more traditional methodology, and potentially serves as a source of triangulation with their practitioners' results. In my study, quantification and automatization are performed through CL tools.

CL

CL has developed what are by now traditional tools to research (large) collections of texts (or corpora), such as wordlists, keywords, concordances, collocations and clusters. More often than not, these traditional tools have been used in synchronic studies. However, they are also being employed in diachronic work, in general, and in modern diachronic approaches (of interest here), in particular. In the latter case, tools are used to examine variation and change 'over comparatively brief periods of modern times' (Partington et al., 2013, p. 265).

CL, then, has been productive in the analysis of differences among corpora. However, increasing interest is now being paid to descriptive work on similarities 'over comparatively brief periods of modern times' (Partington et al., 2013, p. 265). And in order to research this, it introduces innovative concepts and working protocols, such as detailed consistency relations, key keywords (and associates), lockwords, c-collocates and s-collocates (see Partington et al., 2013). These protocols have been there for some time but are clearly underused within Translation Studies (TS). Of all of them, the first two sets of instruments — detailed consistency relations and key keywords — are of special interest to this research (see Section 3).

We are now ready to perform an (MD-)CADS study of original and translated interventions at the EP plenary (from 2005, 2008, and 2010). Prior to the analysis, however, it is apt to describe the European Comparable and Parallel Corpus (ECPC) archive, upon which my analysis is carried out, and to briefly acknowledge the concordancer used in the investigation.

2. The ECPC Archive

The ECPC archive, compiled at the Universitat Jaume I (Spain),[2] is a collection of corpora containing speeches from 1996 to 2011 delivered in English and Spanish at three European Chambers: the EP, the British House of Commons (HC) and the Spanish Congreso de los Diputados (CD). The EP_en corpus, which is used for the present research, consists of speeches from the EP proceedings in English. These sessions are published in the *Official Journal of the European Union* and are downloadable from the EP's website.[3] Within this corpus, there are two subcorpora: EP_en_ST (when English is the original language of the speeches) and EP_en_TT (when the speeches are translated from other languages into English).

This article focuses on original speeches in English from 2005 (928,920 words), 2008 (1,004,733 words) and 2010 (992,094 words). These are compared with translated English for the same years: 2005 (2,304,977 words), 2008 (2,693,074 words) and 2010 (2,460,704 words). As can be seen from the figures, the number of translated English tokens (or running words) is double that of the original English in that combined period, confirming the enormous informative load conveyed via translation within the EP setting.

The explanation for the three-year choice lies in the fact that, according to Partington et al. (2013, p. 321), 'the comparison of the snapshots can reveal even small changes. But to infer with any certainty whether the change was gradual or sudden, the observer needs more than two data points'. For their part, the specific three years chosen for study (i.e. 2005, 2008, and 2010) are relevant to the recent history of the EU, in general, and of the EP, in particular. In this period, the EU has gone from an economic and institutional bonanza to a deep political and financial crisis. This spectacular transition provides an interesting setting within which to examine (original and translated) language.

For the sake of contrast, the article occasionally refers to HC speeches from 2005 (8,597,630 words), 2008 (10,301,549 words) and 2010 (19,300,496 words). In the same way that three temporal data points prove useful to assess the nature of similarity (and change) within the EP, these three *spatial* data points have the same function. Similarity (and difference) is often best seen when compared to a third, external party. For example, by comparing corpus word totals the difference between the original and translated EP English is put into perspective. There is undoubtedly an enormous amount of translated English in EP proceedings. However, the quantitative difference between the (original) Englishes from the EP and the HC is much more prominent.

With regard to these subcorpora, let it be made clear that, overall, this article aims to examine the EP's official *(re)presentations* of speech meaning rather than speech meaning *per se*. The EP_en (and HC) corpus encapsulates some of the primary official images/representations of the EP (and HC) debates in English; as such, it is an important source of linguistic and ideological material.

---

[2] The ECPC archive is freely available to the academic community at http://ecpc.xtrad.uji.es/glossa/html/index.php?corpus=ecpc. Notice that the website is currently under construction.

[3] http://www.europarl.europa.eu/plenary/en/debates-video.html

The main concordancer that is used to query the subcorpora introduced above is monolingual WordSmith Tools 6.0 (Scott 2012), which carries out the three stages of analysis outlined in the following section.

3. The Analysis

As proposed above, the first, immediate aim of this research is to study original and translated EP speeches in English from 2005, 2008 and 2010 (as compiled in the ECPC archive) along the translational and diachronic axes. My ultimate aim is to answer the question: What is kept and what is lost without the written translation of EP proceedings (which stopped being produced in July 2011)? To achieve both aims (and contrary to standard practice in TS), this study does not only focus on ST-TT difference; it also examines corpus similarities. In essence, the present paper revolves around ECPC key keywords, which are generated first (Stage 1). With them, I establish detailed consistency relations (see Stage 2). This gives us a first, quantitative image of the overlapping of idiosyncratic content in the corpora. Afterwards, I identify, group and sort the top 50 key keywords in all EP_en_ST, EP_en_TT and HC subcorpora (Stage 3). In this study (and following DHA protocols), then, I commute from means of realization through strategies to content in order to examine procedures at the linguistic level before July 2011, when both the English original and translated EP proceedings stood side by side. After each descriptive analysis in Stage 3, I briefly pause to discuss the findings.

Stage 1: Generating key keywords

Key keywords are words that are 'key in many texts' (WordSmith Tools 6.0 Online help). They are useful for the present study because they allow us to place the proceedings speech at the centre of investigation, unlike 'plain', 'traditional' keywords, which have the whole corpus as their main unit of analysis.

The process of generating key keywords is quite easy. Using WordSmith Tools 6.0. batch facilities, the analyst compares each corpus text (in this case, speech) with a reference corpus (in this case, the British National Corpus, BNC), identifying those words that are significantly more frequent in that speech than in the reference corpus (i.e. keywords).

With regard to the validity of the BNC for this research, Scott (2009) shows that almost any corpus can be used as reference with productive results. At any rate, the 100-million-word BNC is not just any corpus, but an undoubtedly well-established general corpus of standard British English from the second half of the twentieth century. Due to these features, the BNC is likely to surface, not sink, differences/idiosyncrasy with the ECPC (the main purpose when one uses a reference corpus).

WordSmith Tools 6.0. builds a database of these idiosyncratic terms, which are displayed according to the number of texts where they appear and frequency of

appearance. For this study, I have generated key keyword databases of all subcorpora under study. These include three databases for EP_en_ST (i.e. ST_2005, ST_2008, and ST_2010) and three databases for EP_en_TT (i.e. TT_2005, TT_2008, and TT_2010). They also include two additional databases for all EP_en_ST speeches combined (i.e. ST_ 05_08_10) and for all EP_en_TT speeches combined (i.e. TT_05_08_10). For the sake of comparability, key keyword databases for the HC have been generated along the same lines. The Log Likelihood for all these databases is 0.001 (a rigorous cut-off point to guarantee statistical significance). These databases are the raw material for which I calculated detailed consistency relations and upon which I performed analyses with the top 50 key keywords in all the subcorpora.

Stage 2: Translational and diachronic similarity through detailed consistency relations

Detailed consistency relations operate upon the joint frequency and the type counts of two (or more) corpora. In WordSmith Tools 6.0, they are measured by the Dice Coefficient (see Thada & Jaglan, 2013), which ranges between 0 and 1. By multiplying this coefficient by 100, the percentage of overlap between the corpora is obtained. This is therefore a quick and efficient procedure to measure the overall area of corpus similarity. Depending on whether the detailed consistency relations are calculated for wordlists, keywords or key keywords, different sorts of overlapping are measured. Detailed consistency relations of key keywords measure the percentage of the overlapping idiosyncratic area.

Detailed consistency relations between and among EP_en_ST subcorpora, EP_en_TT subcorpora and HC subcorpora may provide preliminary data to help us hypothesize the linguistic texture that might be kept and lost with the disappearance of EP translation. In addition, detailed consistency relations may also function as indicators of the place the EP_en_ST English occupies in relation to EP_en_TT and HC English. If ST speeches from the EP are closer to the TT English from the Euro-Chamber than to the original English from the HC, this may provide some quantifiable substantiation for the hypothesis that advocates the existence of Euro-jargon, which has already been aptly put forward by scholars, but mainly on the basis of introspection and/or experience (see Koskinen, 2008, p. 43). This procedure may then serve as a complementary perspective to triangulate qualitative work.

A first global comparison along the translational axis, between ST_05_08_10 and TT_05_08_10, shows that there is a 53% key keyword overlap between both sets of corpora. This means that EP ST and TT Englishes share over half of the most characteristic linguistic features and topics discussed at the European Parliament. This is a high proportion, especially if we compare it with what happens when both sets of corpora are examined against the English from the HC. In this case, there is only a 44% overlap between HC_05_08_10 and ST_05_08_10 and a 45% overlap between HC_05_08_10 and TT_05_08_10. It is true that the HC is a much larger corpus than the other two, and it could be argued that it is not therefore quantitatively comparable with

EP material. However, it is also true that the three subcorpora are qualitatively comparable since they contain English production for the very same period of time.

At this stage of analysis, two immediate conclusions can be drawn. Firstly, with the elimination of translation, 53% of idiosyncratic linguistic material is kept. It follows that 47% is lost (or unaccounted for). This is indeed a large quantity for citizens to do without. When interested in EP proceedings, many Europeans do not receive nearly as much information as that at hand. We could hypothesize that in not having access to this information, many topics and positions on them (most probably those concerning other European countries) remain unpublished, and access to them becomes less direct. This could mean that the elimination of translation has had an impact on the transparency that the EP has promised to abide by as well as the hybridity they are so keen on defending. If 'the Other' is silenced to such a great extent, this could eventually have certain consequences for the European sense of identity (that it is proving so difficult to build). These are only hypotheses; but they are hypotheses supported by substantial quantitative data that seem to be worth triangulating by qualitative/ethnographic research, such as undertaken by Abélès (2004), Koskinen (2008) and Wodak (2011). Secondly, there are quantitative data to support that EP ST and TT Englishes are considerably closer to each other than to HC English (by almost 10 percentage points), which may support the existence of a Euro-jargon that transcends the translational divide.

If I now study the diachronic overlap inside all three subcorpora, it becomes apparent that results are particularly stable for the EP chamber regarding both original and translated English. ST speech productions from 2005 and 2008 share 67% of the same idiosyncratic texture and content area. From 2008 and 2010, the overlap is again 67%. Comparing 2005 and 2010, the overlap is 66%. TT speech production overlaps at a greater, though slightly less stable, pace. TT English texture and context in 2005 and 2008 coincide at 70%; in 2008 and 2010, the shared key keyword terrain is 71%, and between 2005 and 2010, it is 67%. The HC results serve to confirm that the EP shared terrain from 2005 through 2008 to 2010 is particularly stable: in the HC, 65% of key keywords coincide in 2005 and 2008; the figure rises markedly to 75% when the years under study are 2008 and 2010, and then decreases considerably to 59% when the measured overlap is between 2005 and 2010.

So what can we say is kept or lost without translation after this internal, diachronic examination? At first sight, not much seems to be lost. Diachronic homogeneity appears to be a constant in the EP regardless of whether speeches are published in the original or translated English. In fact, the EP as a whole seems to be less exposed to the impact of time than the HC. If anything, the figures suggest that the disappearance of translation seems to have protected this stable homogeneity, but only minimally.

Stage 3: Translational and diachronic key keyword examination: from texture through strategies to content

Now that I have measured key keyword overlapping along the translational and diachronic axes, it is time to take a closer look at the specific key keywords obtained in

the study. As in the previous section, I will first display an overall translational picture, after which I will restrict myself to sketching three internal, diachronic snapshots by way of illustration.

The overall picture mainly proceeds from the top 50 key keywords in ST_05_08_10, TT_05_08_10. As part of the analysis, I classified the top 50 key keywords in each set within grammar groups and sorted them further according to meaning proximity in order to spot patterns of strategic use and subsequent content. Notice that by focusing on the corpora's top 50 key keywords, I am placing emphasis on the most noticeable characteristic texture, strategies, and content. For the sake of comparability, these key keywords are presented in alphabetical order. The results of this classificatory grouping and sorting practice are displayed in Figure 1 (ST_05_08_10) and Figure 2 (TT_05_08_10). Those key keywords that are different in both corpora are written in UPPERCASE.

Pre-noun	Pronouns	Adjectives	Nouns	Verbs	Links	Others
Madam Mr This The	I OUR We	Economic European Financial Human Important	Commission Council Countries EU Europe Member Parliament Presidency States Union Citizens Commissioner President Rapporteur Debate Directive ISSUE ISSUES Policy Proposal Report Treaty Agreement Cooperation GOVERNANCE Rights Transparency Lisbon	AM Is Have NEED Should SUPPORT Will	That	S VERY

Figure 1: Top 50 key keywords in ST_05_08_10

Pre-noun	Pro-nouns	Adjectives	Nouns	Verbs	Links	Others
GENTLEMEN Madam Mr This The	We I	Economic European Financial Human Important	Commission Council Countries EU Europe Member Parliament Presidency States Union Citizens Commissioner President Rapporteur RAPPORT-EURS Debate Directive MEASURES Proposal Report STRATEGY Agreement Cooperation CRISIS ENERGY Rights Transparency Lisbon	BELIEVE CONGRAT-ULATE Is MUST Should Will	ALSO That	EUR S

Figure 2: Top 50 key keywords in TT_05_08_10

As seen, we find that ST and TT proceedings speeches (for 2005, 2008 and 2010) share very similar (and almost identical) means of realization and strategic patterns.

As far as nomination (see Section 1) and participants (conveyed mainly via nouns) are concerned, both sets of speeches build their discourses around:

EU institutions: Commission, Council, EU, Europe, Parliament, Presidency, Union

National Institutions: Countries, Members, States

EU agents (at different level): Citizens, Commissioner, President, Rapporteur

EU (preferably EP-related) instruments at different levels of specificity: Debate, Directive, Policy, Proposal, Report, Treaty

Areas of application: Agreement, Cooperation, Rights, Transparency.

These texture and strategic patterns point toward the focal points discussed in (ST and TT) EP speeches. The parliamentary debate, therefore, stays mainly at the European and national levels ('regional' matters, for example, are not particularly characteristic).

The few differences between ST and TT key key nouns are found in categories that I have labelled as 'instruments' and 'areas of application'. While ST speeches refer to the abstract 'Issue(s)', the TTs choose the apparently more technical 'Measure' and 'Strategy'. While ST speeches focus on 'Governance' in an especially prominent way, TT speeches put the attention on 'Crisis' and 'Energy'. Although these differences may not be seen as particularly striking from a quantitative perspective, they may have some qualitative impact upon the preferred content. According to the commonly used WordReference Online, 'Issue' may be defined as 'a point in question or a matter that is in dispute or may be argued about because of its importance'. For its part, 'Measures' (in the plural, the way it is prominently used in the EP) describes 'actions to achieve an end.' Thus, in this particular case, the TT speeches bring the content to the material/action terrain (in Hallidayan terms). It could even be argued that 'Measures' can be regarded as more institutional than 'Issue(s)' if we take into account that collocations such as 'coercive measures' (implemented by controlling institutions) appear to be standard in dictionaries such as the English Collins-Cobuild Online.[4] Also, 'Strategy' is charged with high doses of institutional content, as Eurovoc shows,[5] when it is particularly associated with European Union law, European construction, economic policy, employment and industrial structures and policy. Finally, apart from the obvious fact that 'Governance' refers to different content than 'Crisis' and 'Energy', it seems particularly interesting to note that translated speeches make up such an important share of the content around the financial and economic crisis that hit Europe in 2008.

Regarding nomination strategies, the main classes of (EU, national) participants remain unchanged without translation. However, it is the details that are altered slightly when the EP_en_TT subcorpora are not published. The fact that some outstanding TT key keywords are more institutional than the ST preferred selection seems almost negligible if we stop to ponder that much of the discussion about the 'Crisis' is published in translated English. If we take into consideration the enormous havoc that the financial and economic crisis has caused in Europe since 2008, then the non-

---

[4] http://dictionary.reverso.net/english-cobuild/coercive%20measures
[5] http://eurovoc.europa.eu/drupal/?q=search&text=strategy&cl=en&page=1

translation of elements conveying potential points of views and solutions to this scenario can certainly be seen as a great loss.

According to DHA, nomination strategies also include the discursive construction of processes (usually via the use of verbs). Judging from key key verbs, differences between the STs and TTs seem to be greater in this respect. While ST content links institutions, agents, instruments and areas of application via relational verbs in the first and third persons ('Am', 'Is'), TTs prominently resort to third person relational form ('Is'). This result may lead us to the interpretation that, without TTs, some original doses of dispassionate, relational meaning ('It is …') are still kept, together with STs' more direct and engaged form of relational style ('I am …'). Nevertheless, much of TTs' detachment seems to be lost. As with the other results presented in this section, this indeed has implications that are worth pursuing in future research.

There are also differences regarding verbs of saying and thinking. These are vehicles of strategies of intensification and mitigation, which diverge between EP_en_ST and EP_en_TT on various accounts. These data suggest that there are higher doses of pragmatic (often mitigating) content in TT than in ST speeches, through introductions such as 'I congratulate' and 'I believe'. Additionally, EP translated English makes very prominent use of 'Must'. This modal is used with deontic content in order to express 'strong obligation' (Swan 2010, 327) or even 'prohibition' (Swan 2010, 328) when in the negative. It is also used with epistemic content, to convey 'complete certainty (positive or negative)' (Swan 2010, 327). In this sense, 'Must' is especially 'monoglossic' (Munday, 2012) because the speaker does not allow much margin for receivers to interact with the text content. Finally, STs opt for 'Very' as one of the top 50 key keywords. Elsewhere (Calzada Pérez, forthcoming), it has been shown that in EP_en_ST, this adverb is particularly employed to intensify appreciation in the third person singular (e.g. 'It is very important') or the first person singular (e.g. 'I very much regret').

Pausing for discussion, one might realize that intensification and mitigation profiles are rather different between the EP corpora under study and might be altered through the disappearance of translation. Possibly the most interesting of these 'losses' is the prominence of TT 'Must', its strong and 'monoglossic' modality of either a deontic or epistemic kind. This pattern seems to contravene the current behaviour of original English (at least in the UK), where, according to Baker (2010, p. 66), strong modality is in clear decline. Drawing on Leech, Baker attributes this to 'a number of trends in English, including democratisation' or to 'Fairclough's concepts of personalisation (1989) and conversationalisation of public discourse (1994)' (Baker 2010, p. 67), often initiated in American English.

Predication strategies, built around adjectives (among other features), may be used to complement previously discussed content, especially that conveyed through nomination. The data show that the top 50 key key adjectives expressing idiosyncratic predication coincide entirely in both STs and TTs (i.e. 'European', 'Economic', 'Financial', 'Human', and 'Important'). With these data, it seems logical to argue that, regardless of whether speeches are original or translated, they prominently discuss content associated with 'European' 'Issues/Measures' of an 'Economic' or 'Financial'

nature, while putting forward a defence of 'Human' 'Rights'. It would also follow that, on this occasion, not much is lost, in essence, without translation.

Perspectivation, in turn, adds to the global content by filtering it via point of view, expressed through the selection of pronouns (among other ways). In this case, both STs and TTs use 'I' and 'We' to achieve this function. According to Wodak (2011, p. 99), this is a characteristic behaviour of MEPs, which she explains in the following way:

> In contrast to the European Commission officials who tended to speak of themselves in terms of 'we' […] the MEPs constructed and performed numerous identities, both professional and personal.

However, what Wodak does not comment on is that STs also make use of prominent indirect perspectivation through the possessive pronoun 'Our'. Therefore, what is kept in this case without translation is the expression of MEPs' multiple identities. What changes is the degree of direct/indirect filters.

Finally, space constraints prevent us from discussing the content emanating from the connection between key keywords to *topoi* and fallacies, since this would require a lengthy discussion, with data provided by other CL tools (such as concordances, collocations and clusters) beyond the remit of the present article. Nevertheless, the appearance of 'Also' among the top 50 key keywords in EP_en_TT points to a stronger input of argumentative cohesion. It is to be highlighted that 'Also' marks a descriptive transition between arguments, without assigning attribution, finality, criticism or any other kinds of nuances to the connected arguments. Without translation, explicitation may seem to lose some strength.

The incorporation of data from an external party (i.e., HC speeches from 2005, 2008, and 2010) would help put into perspective the results I have briefly commented on. However, space constraints prevent us from examining these data. Suffice it to say that HC figures would come to confirm the close similarities in texture, strategies, and content that exist between EP_en_ST and EP_en_TT. This would be a second confirmation (the first having been described in the previous section) that EP original English is closer to EP translated English than to HC original English, providing further evidence of Euro-jargon. The HC data would also show that the British Chamber shares more texture, strategies and content with EP_en_ST than with EP_en_TT. To conclude the translational examination of corpus key keywords, it must be borne in mind, however, that in spite EP ST and TT proximity, the data discussed in this section show that there are some (quantifiable) textural differences between original and translated speeches that alter strategies and global content.

A final, internal, diachronic examination of corpus key keywords, drawing on a simplified version of Baker (2011), adds complementary (nuanced) information to the discussion. Due to space constraints, I can only sketch three different types of examples to conclude this key keyword study.

If we focus on texture conducive to the conveyance of nomination strategies, we see that 'Crisis' is only among the top 50 key keywords in EP_en_TT. This does not, of course, mean that the term is not used in ST speeches; it means that it is especially

prominent in TTs vis-à-vis the British National Corpus of English. In spite of this initial global difference, the diachronic profile of the word in the ST subcorpus (Relative Frequency per million words of 109 in 2005; 374 in 2008; and 781 in 2010) shows similarities to the profile of the TT subcorpus (Relative Frequency per million words of 183 in 2005; 674 in 2008; and 1042 in 2010). Both profiles are of an overall ascending nature and both experience a steep rise (as measured by the Coefficient of Variance or CV--i.e. the standard deviation as a percentage of the means-- which is 60.14 for STs; 75.13 for TTs). Incidentally, the word's profile is very different in the HC (Relative Frequency per million words of 26 in 2005; 112 for 2008; 65 for 2010; and CV of 63.64), where the trend is erratic (i.e. neither ascending nor descending).

If we now focus on texture conducive to predication strategies, we find an example of a very different diachronic profile. The adjective 'economic' is one of the top 50 key keywords in EP_en_ST and EP_en_TT. Yet the similarities stop here. The diachronic profile in the STs is now erratic (Relative Frequency per million words of 723 in 2005; 514 in 2008; 1184 for 2010; and CV of 40.97). By contrast, there is a clear ascending trend in the TT subcorpus (Relative Frequency per million words of 923 in 2005; 991 in 2008; 1380 for 2010; and CV of 58.29). Incidentally, the HC trend differs from both EP_en_ST and EP_en_TT. Contrary to what happens in the EP, 'Economic' is not among the top 50 key keywords in the British House. Its time progression is also very different (Relative Frequency per million words of 121 in 2005; 272 in 2008; 213 for 2010; and CV of 37.67).

The final diachronic profile described in this section represents an entirely different case. It refers to the pronoun 'I', used as a key keyword in all three corpora under study. In EP STs (Relative Frequency per million words of 11195 in 2005; 9599 in 2008; 10847 for 2010; and CV of 7.08) and TTs (Relative Frequency per million words of 10626 in 2005; 11160 in 2008; 10473 for 2010; and CV of 54.06), 'I' behaves erratically. However, what is particularly striking here is that, in the ST corpus, the very low CV indicates that 'I' is one of the most stable terms in the corpus, whereas in TTs, the high CV reveals that the word is still experiencing significant ups and downs. Incidentally, the HC stands in between with respect to this key keyword (with Relative Frequency per million words of 6623 in 2005; 9848 in 2008; 6947 for 2010; and CV of 22.75).

These final diachronic data serve to emphasize the importance of a modern diachronic approach to translation. Studies of the global picture of corpora are indeed powerful sources of information (as seen above); however, internal diachronic examinations can provide further clues as to how this overall picture comes into being.

Conclusions

A macro-level examination of original and translated EP debates from 2005, 2008, and 2010 suggests that, without translation, much of the texture, strategies, and content delivered at the Euro-Chamber is nonetheless retained. However, the same examination may lead to the conclusion that, without translation, much may also be lost. This article

provides preliminary data allowing us to produce a first quantification of what remains (53%) and what has been lost (47%). Diachronic (in)stability is also measured. In addition, the article points to potential areas where content might be kept or lost. Among the information kept, nomination participants and predication strategies excel, together with the multiple identifications of MEPs. Among the losses, nuances regarding institutional affiliation, detachment, indirectness and patterns of mitigation and intensification might be highlighted. It must be noted that nuances are not to be easily disregarded, especially in political communication, where it is the detail that brings about the most important breakthroughs (or failures). Also, the present article briefly illustrates three cases of diachronic progress towards the status quo. This progress is again partly similar and partly different between original and translated speeches.

Nevertheless, this study is only the beginning of what will hopefully become a wider discussion. Further research is required to confirm or challenge the data presented here, and DHA's final connection (between content and context) is anticipated to derive from the qualitative (ethnographic) research of Abélès (2004), Koskinen (2008) and Wodak (2011).

Funding

References

Abélès, M. (2004). Identity and borders: An anthropological approach to EU

institutions. *Twenty-First Century Papers: Online Working Papers from the*

Center for the 21st Century Studies, *4,* 1–26.

Baker, P. (2006). *Using corpora in discourse analysis*. London: Continuum.

Baker, P. (2010). *Sociolinguistics and corpus linguistics*. Edinburgh: Edinburgh

University Press.

Baker, P. (2011). Time may change, but we will always have money: diachronic

variation in recent British English. *Journal of English Linguistics*, 39(1), 65–88.

doi: 10.1177/0075424210368368.

Baker, P., Gabrielatos, C., Khosravinik, M., Krzyzanowski, M., McEnery, T., &
    Wodak, R. (2008). A useful methodological synergy? Combining critical
    discourse analysis and corpus linguistics to examine discourses of refugees and
    asylum seekers in the UK press. *Discourse & Society*, 19(3), 273–206. doi:
    10.1177/0957926508088962.

Baker, P., & McEnery, A. (2005). A corpus-based approach to discourses of refugees
    and asylum seekers in UN and newspaper texts. *Journal of Language and
    Politics*, 4(2), 197–226. doi: 10.1075/jlp.4.2.04bak

Barenreuter, C. (2005). Researching the European public sphere and its political
    functions: A proposal. In T. Nesbit & J. Steinberg (Eds.), *Freedom, justice and
    identity*. Vienna: IWM Junior Visiting Fellows' Conferences.

Bauman, Z. (2004). *Europe: An unfinished adventure*. Cambridge; Malden, MA: Polity.

Bayley (Ed.). *Cross-cultural perspectives on parliamentary discourse* (pp. 87–109).
    Amsterdam/Philadelphia: John Benjamins.

Calzada Pérez, M. (forthcoming). Corpus-based methods for Comparative Translation
    and Interpreting Studies: Mapping differences and similarities with traditional
    and innovative tools. *Translation and Interpreting Studies*, 12(2).

Carpuat, M. (2014). Mixed Language and Code Switching in the Canadian Hansard. In
    *Proceedings of The First Workshop on Computational Approaches to Code
    Switching* (pp. 107–115). Doha: Association for Computational Linguistics.

Dibattista, D. (2004). Legitimising and informative discourse in the Kosovo debates in
    the British House of Commons and the Italian Chamber of Deputies. In P.
    Bayley (Ed.), *Cross-cultural perspectives on parliamentary discourse* (pp. 151–
    184). Amsterdam/Philadelphia: John Benjamins.

Eemeren, F. H. van, & Houtlosser, P. (2006). The case of pragma-dialectics. In S.

    Parsons, N. Maudet, P. Moraitis, & I. Rahwan (Eds.), *Argumentation in multi-*

    agent systems (pp. 1–28). Berlin: Springer.

Garzone, G., & Santulli, F. (2004). What can corpus do for critical discourse analysis?

    In A. Partington, J. Morley, & L. Haarman (Eds.), *Corpora and discourse* (pp.

    351–368). Bern: Peter Lang.

Hardt-Mautner, G. (1995). 'Only connect.' Critical discourse analysis and corpus

    linguistics. In *UCREL Technical Paper 6*. Lancaster: University of Lancaster.

Halliday, M.A.K. (1985). *An introduction to functional grammar*. London: Arnold.

Koller, V., & Mautner, G. (2004). Computer applications in critical discourse analysis.

    In C. Coffin, A. Hewings, & K. O'Halloran (Eds.), *Applying English grammar:*

    functional and corpus approaches (pp. 216–228). London: Arnold.

Koskinen, K. (2008). *Translating institutions: an ethnographic study of EU translation*.

    Manchester: St. Jerome.

Krzyżanowski, M. (2005). Analyzing European Union discourse. Theories and

    applications. In R. Wodak & P. A. Chilton (Eds.), *A new agenda in (critical)*

    discourse analysis. Theory, methodology and interdisciplinarity (pp. 137–163).

    Amsterdam/Philadelphia: John Benjamins.

Mautner, G. (2009a). Checks and balances: How corpus linguistics can contribute to

    CDA. In R. Wodak and M. Meyer (Eds.), *Methods of critical discourse analysis.*

    (pp. 122–143). London: Sage.

Mautner, G. (2009b). Corpora and critical discourse analysis. In P. Baker (ed.)*,*

    Contemporary corpus linguistics (pp. 32–46). London: Continuum.

Mehan, H. (1997). The discourse of the illegal immigration debate: A case study in the politics of representation. *Discourse & Society*, 8(2), 249–270. doi: 10.1177/0957926597008002006.

Munday, J. (2012). *Evaluation in translation: Critical points of translator decision-making*. Milton Park: Routledge.

O'Halloran, K., & Coffin, C. (2004). Checking overinterpretation and underinterpretation: Help from corpora in critical linguistics. In C. Coffin, A. Hewings, and K.A. O'Halloran, (Eds.), *Applying English grammar: Functional and corpus approaches.* (pp. 275–297). London: Hodder Arnold.

Orpin, D. (2005). Corpus linguistics and critical discourse analysis: Examining the ideology of sleaze. *International Journal of Corpus Linguistics*, 10(1), 37–61. doi: 10.1177/0957926597008002006.

Partington, A. (2004). Corpora and discourse: A most congruous beast. In A. Partington, J. Morley, & L. Haarman (Eds.), *Corpora and discourse* (pp. 11–20). Bern: Peter Lang.

Partington, A., Duguid, A., & Taylor, C. (Eds.). (2013). *Patterns and meanings in discourse: theory and practice in corpus-assisted discourse studies (CADS)*. Amsterdam/Philadelphia: John Benjamins.

Scott, M. (2009). 'In search of a bad reference corpus'. In D. Archer (Ed.), *What's in a Word-list? Investigating word frequency and keyword extraction* (pp. 79-92). London: Ashgate.

Scott, M. (2012) *WordSmith Tools version 6*. Stroud: Lexical Analysis Software.

Stubbs, M. (1996). *Text and corpus analysis: Computer-assisted studies of language and culture*. Oxford: Blackwell.

Thada, V., & Jaglan, V. (2013). Comparison of Jaccard, Dice, Cosine similarity coefficient to find best fitness value for web retrieved documents using genetic algorithm. *International Journal of Innovations in Engineering and Technology*, 2(4), 202-205.

Wodak, R. (2011). *The discourse of politics in action: politics as usual*. New York: Palgrave Macmillan.

Wodak, R., De Cillia, R., Reisigl, M., & Liebhart, K. (1999). *The discursive construction of national identity*. Edinburgh: Edinburgh University Press.

Wodak, R., & Meyer, M. (2009). Critical discourse analysis: History, agenda, theory and methodology. In R. Wodak & M. Meyer (Eds.), *Methods of critical discourse analysis* (pp. 1–33). London: SAGE.