Meta
Journal des traducteurs
Translators' Journal

Simplified or not Simplified? The Different Guises of Mediated English at the European Parliament

Adriano Ferraresi, Silvia Bernardini, Maja Miličević Petrović et Marie-Aude Lefer

Volume 63, numéro 3, décembre 2018

Traductologie de corpus : 20 ans après

URI : https://id.erudit.org/iderudit/1060170ar
DOI : https://doi.org/10.7202/1060170ar

Aller au sommaire du numéro

Éditeur(s)

Les Presses de l'Université de Montréal

ISSN

0026-0452 (imprimé)
1492-1421 (numérique)

Découvrir la revue

Citer cet article

Ferraresi, A., Bernardini, S., Petrović, M. M. & Lefer, M.-A. (2018). Simplified or not Simplified? The Different Guises of Mediated English at the European Parliament. *Meta*, *63*(3), 717–738. https://doi.org/10.7202/1060170ar

Résumé de l'article

Dans cet article, nous proposons un cadre méthodologique pour l'analyse comparative de l'interprétation simultanée et de la traduction écrite, proposition que nous illustrons par une étude de la simplification lexicale dans le cadre de modalités de production linguistique (langue orale vs écrite) et de combinaisons de langues différentes. Notre étude repose sur le corpus intermodal EPTIC, qui comprend les interprétations et les traductions des séances plénières du Parlement européen, alignées avec leurs textes sources, pour deux paires de langues (anglais><italien et anglais><français). L'objectif de notre étude étant de mettre au jour les phénomènes liés à la simplification dans deux modalités de médiation interlinguistique, nous comparons une série de traits lexicaux, tels que la densité lexicale, en anglais traduit et interprété à partir du français et de l'italien, à la fois d'un point de vue comparable monolingue et d'un point de vue intermodal. Les résultats obtenus ne confirment que partiellement l'hypothèse de simplification en langue médiée : la simplification lexicale observée en anglais médié est plus marquée (i) lorsque la langue source est le français, et (ii) dans les interprétations. Nous concluons que la simplification dépend à la fois de la langue source, et, dans une moindre mesure, de la modalité de médiation interlinguistique.

Simplified or not Simplified? The Different Guises of Mediated English at the European Parliament

ADRIANO FERRARESI
Università di Bologna, Forlì, Italy
adriano.ferraresi@unibo.it

SILVIA BERNARDINI
Università di Bologna, Forlì, Italy
silvia.bernardini@unibo.it

MAJA MILIČEVIĆ PETROVIĆ
Univerzitet u Beogradu, Belgrade, Serbia
m.milicevic@fil.bg.ac.rs

MARIE-AUDE LEFER
Université catholique de Louvain, Louvain-la-Neuve, Belgium
marie-aude.lefer@uclouvain.be

RÉSUMÉ

Dans cet article, nous proposons un cadre méthodologique pour l'analyse comparative de l'interprétation simultanée et de la traduction écrite, proposition que nous illustrons par une étude de la simplification lexicale dans le cadre de modalités de production linguistique (langue orale vs écrite) et de combinaisons de langues différentes. Notre étude repose sur le corpus intermodal EPTIC, qui comprend les interprétations et les traductions des séances plénières du Parlement européen, alignées avec leurs textes sources, pour deux paires de langues (anglais><italien et anglais><français). L'objectif de notre étude étant de mettre au jour les phénomènes liés à la simplification dans deux modalités de médiation interlinguistique, nous comparons une série de traits lexicaux, tels que la densité lexicale, en anglais traduit et interprété à partir du français et de l'italien, à la fois d'un point de vue comparable monolingue et d'un point de vue intermodal. Les résultats obtenus ne confirment que partiellement l'hypothèse de simplification en langue médiée : la simplification lexicale observée en anglais médié est plus marquée (i) lorsque la langue source est le français, et (ii) dans les interprétations. Nous concluons que la simplification dépend à la fois de la langue source, et, dans une moindre mesure, de la modalité de médiation interlinguistique.

ABSTRACT

In this article we describe a framework for the corpus-based comparative investigation of interpreting and translation, illustrating it through a study of simplification across different modes of language production and across different language pairs. We rely on EPTIC, a corpus featuring plenary speeches at the European Parliament in their interpreted and translated versions, aligned to each other and to their source texts in English<=>Italian and English<=>French. Aiming to shed light on lexical simplification in different mediation modes, we compare interpretations and translations to each other and to comparable original speeches and their edited written versions. Specifically, we compare lexical features (lexical density, type-token ratio, core vocabulary and list head coverage) in interpreting and translation into English from French and Italian, both in a monolingual comparable perspective and an intermodal perspective. Our results do not unconditionally support the simplification hypothesis: lexical simplification is observed in mediated English, but is found to be greater when the source language is French, and

in interpretations rather than translations. We conclude that this feature is contingent on both the mediation mode and the source languages involved, and that the influence of the latter seems to be stronger than that of the former.

RESUMEN

En este trabajo describimos un marco para la investigación comparativa basada en corpus de la traducción y la interpretación, ilustrándolo a través de un estudio de la simplificación en diferentes modalidades de producción lingüística y diferentes pares de lenguas. Nos basamos en EPTIC, un corpus que incluye discursos de sesiones plenarias del Parlamento Europeo en sus versiones interpretadas y traducidas, alineadas entre ellas y con sus textos originales en inglés<=>italiano e inglés<=>francés. Con el objetivo de arrojar luz sobre la simplificación léxica en diferentes modalidades de mediación, comparamos interpretaciones y traducciones entre ellas y con discursos originales comparables y sus versiones escritas editadas. Más específicamente, comparamos las características lexicales (densidad léxica, *type-token ratio*, *core vocabulary* y *list head coverage*) en interpretación y traducción del francés y del italiano hacia el inglés, tanto desde una perspectiva monolingüe comparable, como desde una perspectiva intermodal. Nuestros resultados no corroboran incondicionalmente la hipótesis de la simplificación: la simplificación léxica se observa en el inglés mediado, pero resulta ser mayor cuando la lengua fuente es el francés, y más en la interpretación que en la traducción. Concluimos que esta característica es contingente tanto en la modalidad de mediación como en las lenguas fuente involucradas, y que la influencia de las segundas parece ser mayor que la de la primera.

MOTS-CLÉS/KEYWORDS/PALABRAS CLAVE

approche basée sur corpus, corpus intermodaux, interprétation, traduction, simplification
corpus-based approach, intermodal corpora, interpreting, translation, simplification
enfoque basado en corpus, corpus intermodal, interpretación, traducción, simplificación

1. Introduction: typical features of translation, twenty years on

Anyone approaching the study of translation through the lens of corpus linguistics is bound to refer back to Baker (1993) as the first major programmatic statement about the theoretical bases, hypotheses, data sources, types of comparisons, and ultimate aims of corpus-based translation studies (CBTS). Yet, when focusing on the first full-fledged application of such a forward-looking programme, the 1998 special issue of *Meta,* edited by Sara Laviosa (Laviosa 1998a), is likely to come to mind. After two decades, we revisit one of the contributions to that volume, namely Laviosa's (1998b) study of lexical regularities in original (non-translated) and translated English narrative prose. This study has had an enormous influence and has inspired investigations of translated language using a host of different corpus setups and methods.

Our focus is on lexical measures that can be interpreted as clues to textual simplification[1] in translation, here intended as a synonym for interlinguistic *mediation*, both oral and written. More specifically, and following Laviosa (1998b), we measure lexical density, type-token ratio, core vocabulary and list head coverage[2] across different subcorpora of EPTIC (Ferraresi and Bernardini 2019), an intermodal corpus of speeches delivered at the European Parliament (EP), in its trilingual (English/French/Italian) version.

Over the years, several researchers have carried out investigations of simplification along the lines of Laviosa (1998b), and some have extended her methodology in different ways. For instance, the set of simplification measures used has been

enlarged, and more computationally sophisticated methods have been adopted (we review this work in Section 2). Our analysis extends the focus of the comparison along two dimensions: first, it includes simultaneous interpretations and written translations; second, it contrasts target texts derived from two different source languages. Other than these two dimensions, the simplification measures and statistical methods were kept as similar as possible to Laviosa's. In this way we ensure comparability of results with her work, as well as with similar later investigations that have focused on different genres, language pairs and modes of text production (also reviewed in Section 2).

By resorting to a shared methodology, our ultimate aim is thus to contribute to assembling "a consistent body of evidence which will progressively refine the initial hypotheses and give rise to further, more precise predictions" (Laviosa 2002: 63-64). We suggest that the bottom-up investigation of hypothesised typical features of translation through numerous focused studies is one of the central ways in which corpus linguistics can contribute to translation studies, providing on-the-ground evidence through which the discipline moves "gradually, and in as controlled a way as possible, from individual instantiations to the culture-specific, to more and more general regularities on higher and higher levels, to generate new, or modified theoretical statements" (Toury 2004: 28).

The paper is structured as follows. Section 2 describes the rationale for our study and reviews relevant previous work on simplification in translation, source language influence, and intermodal corpora. The setup of the corpus is described in Section 3, the methodology used for our study in Section 4, and its results in Section 5. Section 6 attempts to interpret these results, relating them to those obtained in previous corpus-based work on simplification in translation and interpreting, and Section 7 concludes by briefly evaluating implications for research on the typical features of mediated language.

2. Study background

2.1. Simplification as a typical feature of translation: the jury is still out

In her ground-breaking work from the late 1990s (Laviosa 1998b; Laviosa 1998c), Laviosa compared translated fiction and newspaper texts to comparable original texts on the basis of several language-independent measures of simplification:

- lexical density, or the percentage of content to function words;
- core vocabulary, or the proportion of high- to low-frequency words;
- list head coverage, or the proportion of the corpus accounted for by its most frequent lemmas;
- type-token ratio, or the proportion of different words to the total number of words in the corpus;
- sentence length, or the average number of words per sentence.

Some of these measures appeared to be sensitive to text type: for instance, sentence length was found to be significantly lower in translated than original newspaper language, but the opposite was true of translated fiction; others seemed more stable, and potential candidates for the status of "core patterns of lexical use" in translation (Laviosa 1998b).

Multiple studies continue to explore these issues. Xiao and Hu (2015) and Hu (2016) review extensive evidence about Chinese translated from English, confirming Laviosa's results. Lower lexical density, lower type-token ratio and greater use of core vocabulary seem to also characterise translated Chinese, along with other clues to simplification, such as the less frequent use of archaic syntactic structures and the more frequent use of common word clusters (Hu 2016: 170). The authors however warn against drawing hasty conclusions. First, like Laviosa, they point out that analyses of different text types return conflicting results for the same simplification parameters, findings that argue against the universalist hypothesis. And second, they find that translated language also displays features of complexification, for example, greater use of very infrequent and long words.

Along similar lines, Williams (2005) confirms Laviosa's findings with respect to the type-token ratio, lexical density and sentence length of English texts translated from French and published on the website of the Canadian Government. However, French texts translated from English from the same source display opposite features: while English translations are simpler, French translations are more complex than the corresponding original texts. She concludes that results "appear to be related in some way to the characteristics of each language of translation, and to the fact that each has the other as a source language, rather than to the translated or non-translated status" (Williams 2005: 143). Similar results are obtained by Sandrelli and Bendazzoli (2005) in their study of list head coverage in interpreted and original (non-interpreted) EU speeches: while interpretations from Italian and Spanish into English display less lexical variety, the opposite is true of interpretations from English and Spanish into Italian, once more hinting at the central role of the languages involved.

Faced with the problem of ascertaining the meaningfulness of statements about typical features of translation, several researchers have turned to more sophisticated natural language processing techniques, including machine learning. Corpas Pastor, Mitkov, *et al.* (2008) analyse simplification in Spanish medical texts translated from English by professionals and by students, and in Spanish technical texts translated by professionals, using a range of measures, such as lexical density, lexical richness and sentence length. The study finds partial evidence of simplification, particularly in terms of lexical density and richness, but not in terms of sentence length, which is higher in translations. Interestingly, student translations are found to be less simplified than professional ones. In an attempt to find "confirmation in a predictive classification test" (Chesterman 2004: 42), Ilisei, Inkpen, *et al.* (2010) use simplification features (such as word and sentence length) alongside other features of translationese to train a text classifier on the same dataset as Corpas Pastor, Mitkov, *et al.* (2008). They find that simplification features are indeed useful to detect translationese, substantially increasing the accuracy of the classifier. Summing up, it seems that:

> [t]he simplification hypothesis […] is in fact testified by a mixture of evidence and counter evidence. What is more reasonable is probably not to verify or deny the simplification hypothesis in an absolute sense but to consider and compare various factors, supportive or subversive, to reach a more detailed and hence more profound understanding of simplification. (Xiao and Hu 2015: 159)

2.2. Looking at source influence from a target perspective

In the present study, one of our aims is to gauge whether the source language affects simplification in interlingual mediation. Source language influence, sometimes referred to as *interference*, has long been hypothesised to characterise translation (Toury 1995). Yet, the assumption that the source language leaves traces in the target text (an "S" universal in Chesterman's (2004) terms) would be at odds with the "T" universal assumption that translated texts are always simpler than comparable original texts in the target language, regardless of the source language. Our study looks at simplification patterns in English texts translated and interpreted from French and Italian, adopting a target perspective to investigate source language influence. A number of studies have tackled source language effects in similar ways.

Van Halteren (2008) uses text classification methods to classify edited written versions of *Europarl* speeches (so-called verbatim reports) translated from and into English, Dutch, French, Italian and Spanish. Relying on typical word sequences, the classifiers are able to identify the source language with an accuracy above 87%. The distinguishing features are found to be both linguistic and discursive/cultural. Koppel and Ordan (2011) carry out a battery of text classification experiments focusing on English newspaper texts and speeches in *Europarl*, using 300 frequent function words as their features. Overall, they cover eight source languages (including Finnish, Hebrew and Korean). Their results indicate that "both source language and the mere fact of being translated play a crucial role in the makeup of a translated text" (Koppel and Ordan 2011: 1324-1325). While "translations from similar source languages are different from non-translated texts in similar ways" (Koppel and Ordan 2011: 1325), certain features, such as the overrepresentation of cohesive adverbs, are observed in translated texts regardless of the source languages.

Text classification methods such as those just reviewed typically rely on shallow lexico-syntactic features. Other studies have attempted to test theoretically-grounded hypotheses about source language effects on target texts. Cappelle (2012) analyses manner of motion verbs in English texts translated from French and German. Since these types of verbs are unusual in French (a verb-framed language), but frequent in German and English (both satellite-framed languages), they are likely to be less frequent in translation from French (than in original English), but not in translation from German. The data confirm this prediction, which finds further support in a methodologically related study by Cappelle and Loock (2017), focusing on phrasal verbs. Similarly, Hareide (2017) compares the use of the Spanish gerund in original Spanish texts, translations from Norwegian (a language with no direct grammatical equivalent) and translations from English (a language with partly overlapping grammatical resources). Both translations from English and translations from Norwegian feature significant over-representation of the gerund with respect to original Spanish, suggesting a tendency for typical features of a language to be over-represented in translation. However, the trait is more prominent in translations from English, pointing to a concurrent source language effect.

Lastly, Kolehmainen and Riionheimo (2016) analyse the Finnish passive in literary translations from German and Estonian and attempt to relate their results to previous work on the use of the same structure in spoken interviews of Finnish migrants in Estonia. While they fail to find confirmation for a shared source language

effect, theirs is a valiant attempt to test hypotheses about features in common between translation and other language contact situations (Kruger and Van Rooy 2016). Indeed, the extensive literature on language contact provides a wealth of suggestions for typical contact-induced properties and ways of testing them (Szmrecsanyi and Kortmann (2009) on simplification and complexification). Investigating their existence and role in translation could significantly extend and deepen the current research agenda in CBTS (operational suggestions are offered by Kranich 2014). Conceptualising translation as a locus of language contact, our study aims to contribute to the research reviewed in this section by investigating the role played by the source language in (different modes of) interlingual mediation.

2.3. Translation, interpreting and the intermodal approach

Our third area of interest is research that tests hypotheses about typical features of interlingual mediation by contrasting *written* translation to *oral* simultaneous interpreting. First discussed by Shlesinger (1998), this testing scenario has received growing attention in recent years, despite the difficulties inherent in setting up adequate corpus resources. These should ideally include authentic interpretations, authentic written translations of the same texts, and authentic, as well as comparable, original speeches (Shlesinger 1998: 488). If building translation corpora is a complex task, building intermodal corpora is even more complex, by several orders of magnitude.

In an attempt to investigate similarities and differences between these modes of interlingual mediation, Shlesinger and Ordan (2012) focus on English-to-Hebrew interpretations and translations in the academic domain. Their results suggest that certain features, found to be typical of written translation in previous studies (Laviosa 1998b), such as simplification, are all the more extreme in interpreting. Hu and Tao (2013) also find evidence of greater levels of explicitation and normalisation in interpreting than in translation, based on a corpus of press conferences and political speeches mediated from English into Chinese.

Several intermodal studies have focused on the treasure throve of EP speeches. Kajzer-Wietrzny (2012) tests hypotheses concerning translation universals in her monolingual comparable intermodal corpus of EP speeches interpreted and translated from Dutch, French, German and Spanish into English. Her findings, consistent with those of Sandrelli and Bendazzoli (2005), suggest that simplification is not consistently found in interpreting, and points to the important role played by the languages involved. Further, in a study of clausal connectives in EP speeches, interpreted and translated from French into English and Dutch, Defrancq, Plevoets, *et al.* (2015) observe a tendency for interpreters to add more connectives than translators do, and thus to explicitate more (their corpus is bilingual, parallel as well as intermodal, thus allowing them to observe interpreter *vs.* translator choices). Finally, Bernardini, Ferraresi, *et al.* (2016) analyse translations and interpretations from Italian into English and from English into Italian in a version of EPTIC containing EP speeches from 2004. They find evidence of greater simplification in interpreted than translated language, but also observe language-specific effects, since English interpreted speeches make more use of frequent words and text-internal repetitions, while Italian ones resort to shorter sentences and greater use of function words (Bernardini, Ferraresi, *et al.* 2016: 80). Since the mediated components of the corpus

are also simpler than the non-mediated components, this study finds support for Shlesinger and Ordan's (2012: 54) portrayal of interpreting "as an extreme case of translation."

The research reviewed in this Section thus suggests that mediated language differs from original, non-mediated language, and that translation differs from interpretation, in predictable ways. However, it also shows that different source languages may reflect upon mediated texts, somewhat contradicting the generalizability of the previous claim. Thanks to the unique design of EPTIC, described in Section 3, this study explores the interaction between these factors.

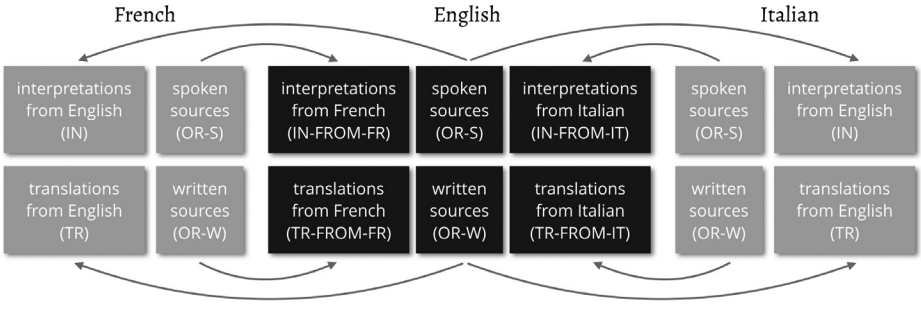
3. EPTIC: the European Parliament Translation and Interpreting Corpus

EPTIC is an intermodal corpus representing two modes of interlinguistic mediation, namely simultaneous interpreting and written translation. Initiated as an offshoot of the *European Parliament Interpreting Corpus* (EPIC; Russo, Bendazzoli, *et al.* 2012), EPTIC is currently being compiled as a joint effort between research teams at several European universities, including the University of Bologna, the University of Belgrade and the University of Louvain. The corpus features four main components: (1) transcripts of speeches delivered at EP plenary sittings (*OR-S* for *original spoken*), where speakers have the right to use the language of their choice, (2) transcripts of the simultaneous interpretations of these speeches (*IN* for *interpreted*), (3) the verbatim reports of the plenary sittings (*OR-W* for *original written*), and (4) the official translations of these verbatim reports (*TR* for *translated*).[3]

The EP has features that make it almost unique as a source of intermodal corpora: first and foremost, the fact that videos of the speeches interpreted in all the working languages (some through relay), verbatim reports and translations (up to 2011) have been made available for easy download from the Parliament website.[4] Furthermore, a wide range of languages are available, used both as source and target languages, and speeches are delivered both in impromptu and prepared delivery modes, in the case of English and a few other languages by native as well as non-native speakers, offering a wealth of dimensions for comparison.

The current version of the corpus features speeches delivered in the first months of 2011 in three languages (English, French and Italian).[5] In the selection of corpus materials, speeches were sampled from a limited number of plenary sittings, so that texts are as comparable as possible in terms of topics covered. At the same time, an attempt was made to maximise the number of different speakers, so as to increase corpus representativeness. EPTIC comprises French and Italian targets for the English sources, and English targets for the French and Italian sources. The resulting corpus setup consists of two comparable and bidirectional sets of subcorpora, one for the English<=>Italian combination, the other for the English<=>French combination, for a total of 14 subcorpora (Figure 1).

Figure 1
The structure of EPTIC



At the time of writing, the overall size of the trilingual version of EPTIC is approximately 280,000 words, with 64 source speeches in English, 65 in French and 68 in Italian (Table 1).[6] This is quite modest compared to other EU-based multilingual resources (for example, *Europarl*), but substantial in terms of intermodal corpora, which at times contain as few as 6 speeches (Shlesinger 2009). Further advantages of EPTIC are its ecological validity and the fact that the sources for interpreting and translation are nearly identical.

Table 1
Trilingual EPTIC (2011 data)

Language	Subcorpus	Number of texts	Total number of running words
English	OR-S	64	21,106
	OR-W	64	20,091
	IN-FROM-FR	65	20,836
	TR-FROM-FR	65	21,641
	IN-FROM-IT	68	16,122
	TR-FROM-IT	68	17,775
French	OR-S	65	23,943
	OR-W	65	23,159
	IN	63	20,558
	TR	64	22,940
Italian	OR-S	68	16,586
	OR-W	68	16,521
	IN	64	16,977
	TR	64	19,665
TOTAL		**915**	**277,920**

As illustrated by the examples in Table 2, the main differences between transcribed speeches and their verbatim reports lie in the deletion of disfluencies, such as false starts and mispronunciations, and in minor lexico-syntactic changes.

TABLE 2
Excerpts of sources (*OR-S* and *OR-W*)

Transcribed speeches (*OR-S*)	Corresponding verbatim reports (*OR-W*)
Today, it's Martin Luther King Day and Martin Luther King said: "A time comes when silence becomes betrayal." Commissioner Füle, that time has long bec- beca- ehm arrived.	Today is Martin Luther King Day and it was Martin Luther King who said that a time comes when silence becomes betrayal. Commissioner Füle, that time has arrived.
But the p-… practicalities of an agreement with a State the size of Congo create all sorts of … all sorts of daunting problems.	Yet the practicalities of an agreement with a State the size of Congo create all sorts of daunting problems.

Metadata related to the speech, the speaker and the interpreter vary according to the subcorpus. In addition to information about the speech (such as date, length, topic, etc.) and its originator (such as name, gender, native speaker status, etc.), collected for all versions, the oral speeches also include metadata about speech duration, speed and mode of delivery (read, impromptu or mixed), and among these, the interpreted speeches also include minimal information about the interpreter (gender and native speaker status).

EPTIC is annotated with part-of-speech and lemma information, using *TreeTagger*,[7] and is multi-aligned at sentence level: the plenary speeches in their interpreted and translated versions in the target languages are aligned to each other and to their source texts. Given the presence of transcribed data, and the loss of information that goes with orthographic transcription, text-to-video alignment is provided, for access to the audio directly from the concordance lines (Ferraresi and Bernardini 2019). The corpus is made available to the public through a *NoSketch Engine* platform (Rychlý 2007) hosted by the University of Bologna.[8]

Thanks to its composite structure, EPTIC affords many types of comparisons: *parallel* (source texts *vs*. target texts), *monolingual comparable* (mediated language *vs*. non-mediated, original language), as well as *intermodal* (simultaneous interpreting *vs*. written translation). The study reported below focuses on English only, taking advantage of the monolingual comparable and intermodal perspectives offered by EPTIC, and leaving the parallel one for future investigations.

4. Method

4.1. Research questions

As mentioned in Section 1, in this study we aim to shed light on lexical simplification in mediated texts by comparing interpretations and translations to each other and to their non-translated and non-interpreted counterparts (original EP speeches and their edited written versions). We attempt to measure the impact of the source language variable, comparing texts mediated from two source languages, French and Italian, into English. The overarching goal is to find out how instances of oral and written, mediated, and original discourse are positioned with respect to each other in terms of lexical simplification, taking the source language into account.

Based on our review of previous work in this area, our research questions are as follows:

- (RQ1) Are English mediated texts in EPTIC lexically simpler than English originals?
- (RQ2) Are English interpreted speeches in EPTIC lexically simpler than translated verbatim reports?
- (RQ3) Are texts mediated from French and Italian in EPTIC equally simplified?

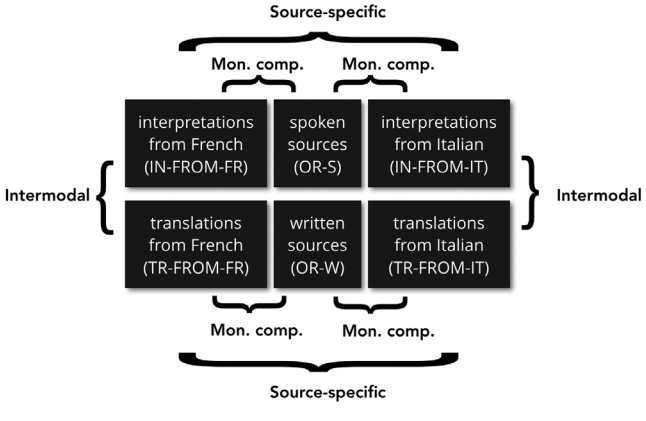
4.2. Monolingual comparable, intermodal and source-specific comparisons

We address RQ1 through monolingual comparable comparisons, carried out by contrasting the interpreted and non-interpreted English subcorpora on the one hand, and the translated and non-translated subcorpora on the other, separately for texts mediated from French and from Italian. An intermodal perspective is adopted for RQ2, where we contrast the interpreted and translated English subcorpora, again separately for the two source languages. Finally, we focus specifically on the source language variable (RQ3), contrasting texts with different source languages, separately for interpreting and translation.

This approach means that the parameters of simplification described in Section 4.3 are tested in eight sets of comparisons (also represented graphically in Figure 2):

1) speeches interpreted from French *vs.* original speeches (monolingual comparable)
2) texts translated from French *vs.* original verbatim reports (monolingual comparable)
3) speeches interpreted from Italian *vs.* original speeches (monolingual comparable)
4) texts translated from Italian *vs.* original verbatim reports (monolingual comparable)
5) interpretations *vs.* translations from French (intermodal)
6) interpretations *vs.* translations from Italian (intermodal)
7) interpretations from French *vs.* Italian (source-specific)
8) translations from French *vs.* Italian (source-specific).

Figure 2
The comparisons carried out in the study



The parallel dimension is explored only indirectly: since the intermodal comparisons apply to interpretations and translations of spoken and written sources that are assumed to be closely related, the source corpus components are checked for possible differences that could preclude intermodal comparisons. If significant differences existed between the source speeches and the corresponding verbatim reports, they would hint at editorial changes in turning the original speeches into

written texts, making the intermodal comparisons uninterpretable with respect to the interpreted/translated distinction. None of the control source comparisons returned significant results, therefore they were disregarded in what follows.

4.3. Simplification parameters and statistical tests

For reasons explained in Section 1, we rely on the method adopted by Laviosa (1998b; 1998c), distancing ourselves only when our research questions or methodological advances in the discipline require it. We examine the four simplification parameters that have proved to be reasonably stable across previous studies: lexical density, type-token ratio, core vocabulary coverage, and list head coverage. Lexical density is taken as a measure of information load, type-token ratio as a measure of lexical diversity in terms of the range of vocabulary used, while core vocabulary coverage and list head coverage reveal patterns of use of very frequent words, providing a different angle on diversity.

Following Laviosa (1998b; 1998c), who in turn refers back to Stubbs (1996: 172), we define lexical density as the proportion of lexical to function words, calculated "by subtracting the number of function words […] from the number of running words (which gives the number of lexical words) and then dividing the result by the number of running words" (Laviosa 1998b, note 3). We count running words using a regular expression that matches sequences of numbers and/or letters, which may include apostrophes and hyphens.[9] Function words are identified based on part-of-speech tags: a list of all tags used in the English subcorpora of EPTIC is obtained, and the tags are manually classified as either lexical (adjective, noun, digit, verb and open-class adverb tags)[10] or functional (all other tags excluding punctuation signs, disfluencies and foreign words).[11]

The type-token ratio is calculated as the ratio of the number of unique word forms (types) to the total number of running word forms (tokens) in the same text.

Core vocabulary coverage refers to the proportion of high-frequency words to low(er)-frequency words, the former being defined as the 200 most frequent words from a large reference corpus. Laviosa's high-frequency word list (also used by Kajzer-Wietrzny 2012) was obtained from the Collins Cobuild Bank of English, while we opted for the list extracted from *ukWaC* (Baroni, Bernardini, *et al.* 2009), which provides more recent (and more accessible) data. The number of occurrences of low-frequency words is calculated by summing up the number of occurrences of each high-frequency word in each EPTIC text and subtracting this figure from the total number of running words for that text.

List head coverage is defined as the proportion of each subcorpus accounted for by the top 100 words in their frequency lists. Unlike core vocabulary coverage, which is determined with respect to an external point of reference, list head coverage is a corpus-internal measure. In the case of EPTIC, whose individual texts can be fairly short, list head status is best defined based on cumulative frequencies at a subcorpus level; Laviosa (1998b) also performed a by-corpus, rather than a by-text analysis with respect to this measure. To obtain the counts of non-list head ("tail") words, we added up the number of occurrences of each list head word in its subcorpus and subtracted this figure from that for the total number of running words in that subcorpus.

To sum up, lexical density, type-token ratio and core vocabulary coverage are computed on a single text basis, whereas list head coverage is calculated on the entire subcorpora. In other words, whenever possible, we use by-text analyses, which allow us to account for individual variation between texts within the different subcorpora (Biber and Jones 2009); by-text analyses also ensure the validity of comparisons across subcorpora of different sizes.

When an analysis is done by text, we perform statistical testing using Mann-Whitney tests to compare the relevant subcorpora; the correlation coefficient *r* is used as an effect size measure. We use non-parametric tests based on the results of preliminary checks on the normality of the distributions (Shapiro-Wilk tests), which revealed that the majority of data is not normally distributed. The Chi-square (χ^2) test is used for by-corpus analyses, accompanied by the phi (φ) coefficient as the measure of effect size. All statistical analyses are performed using the *R* software.[12]

5. Results

5.1. Overview

To provide a general overview of the data, we show the distribution of values for the studied simplification parameters in the different subcorpora. Results of by-text analyses are visualised using boxplots, which show both the central tendency and the spread of the data, while a mosaic plot is used for list head coverage, calculated on the entire subcorpora. In all graphs, subcorpora are grouped by modality, with oral corpora to the left (original, interpreted from French and interpreted from Italian), and written corpora to the right (original, translated from French and translated from Italian). Proportions are expressed as percentages.

Figure 3 shows the results for lexical density, which is highest in texts translated from Italian, and lowest in texts interpreted from French (as evidenced by the median values represented by the thick black lines). Interestingly, in original texts lexical density is higher in oral than in written production. A very similar trend can be seen in the results for the type-token ratio (Figure 4): the highest values are found in texts

Figure 3
Lexical density in original and mediated English

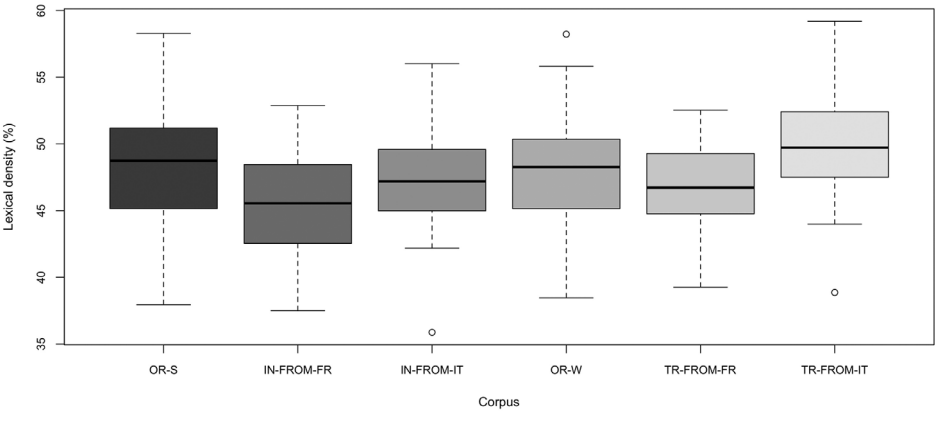
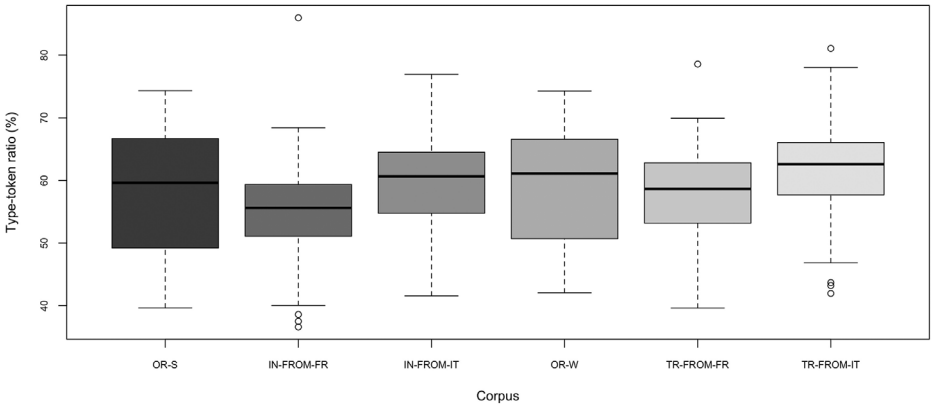


FIGURE 4
Type-token ratio in original and mediated English



translated from Italian and the lowest values in texts interpreted from French. In this case, however, original written texts behave like the mediated ones: written texts display higher values than oral texts. For both translated and interpreted texts, the values for Italian as a source language are also consistently higher than the values for French as a source language.

Overall, the results from above are coherent with those obtained in the analysis of core vocabulary coverage (Figure 5). Texts translated from Italian and interpreted from French are the most and the least lexically diverse ones, that is, they comprise the lowest and highest number of (externally selected) high-frequency words. Texts mediated from Italian show more diversity than texts mediated from French. Among original texts, the oral texts are slightly more diverse than the written ones.

Lastly, list head coverage, shown in Figure 6, seems to be fairly similar across corpora, with interpreted subcorpora being only slightly more repetitive than the rest.

We next move on to report the results of statistical testing for the eight sets of comparisons outlined in Section 4.2.

FIGURE 5
Core vocabulary coverage in original and mediated English

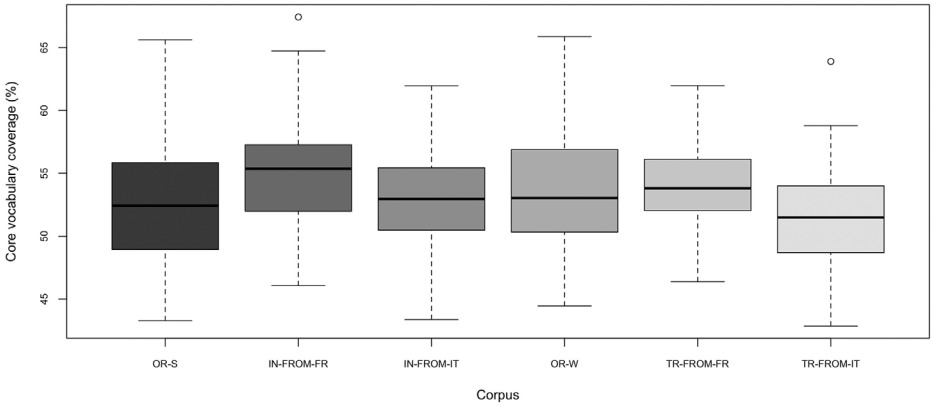
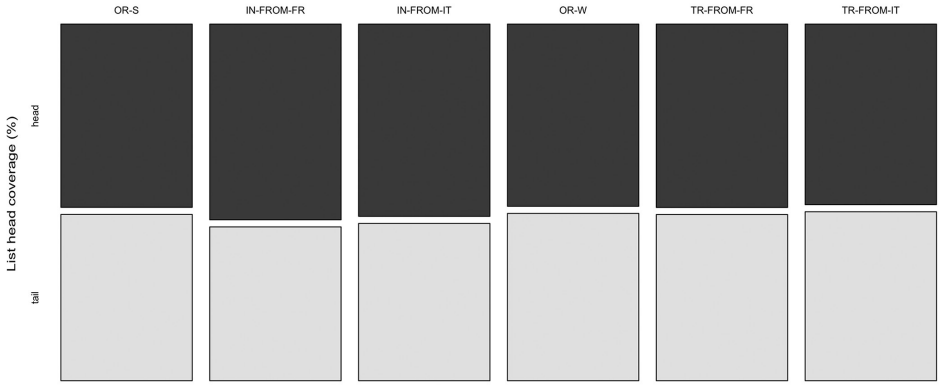


Figure 6
List head coverage in original and mediated English



5.2. Monolingual comparable comparisons

Starting from RQ1 and the monolingual comparable perspective, Tables 3 and 4 show the values for the four studied parameters in the interpreted and translated subcorpora. For each parameter, we present the median values for texts mediated from French and Italian and the values for original texts (note that these values are repeated for the two source languages, as the reference point is the same in both cases). The results of statistical tests are shown in the next-to-last column, with significant differences in bold. The last column summarizes results by reporting the name of the subcorpus in which greater simplification was observed at statistically significant levels. Median values derived from single text measures and results of Mann-Whitney tests are reported for the three by-text comparisons (lexical density, type-token ratio and core vocabulary coverage). The values concerning list head coverage, which are calculated on a by-corpus basis, are single direct measures; the results of χ^2 tests reported next to them are performed on raw frequencies from which these measures are derived.

The results of the statistical tests confirm the observations made in Section 5.1 based on data visualisations. For the parameters of lexical density and core vocabulary coverage, speeches interpreted from French differ significantly from comparable original speeches, while for translations, significant differences are found for texts with Italian as the source language. Speeches interpreted from French are lexically simpler than original English speeches, whereas texts translated from Italian are more complex than original English texts. The effect size is medium in all cases (Field, Miles, *et al*. 2012: 665).

A similar pattern is found for the type-token ratio, but the differences for this parameter do not reach statistical significance. This is probably due to the large amount of variation in the data, especially in the subcorpora of original texts, as evidenced by the height of the boxes in Figure 4.

TABLE 3

Interpreted vs. original spoken English

Simplification parameter	Source language	Results (M,%)		Statistical tests					Simplification in subcorpus
		IN	OR-S	W	r	p	$\chi^2(1)$	φ	
Lexical density	French	45.54	48.74	**2738.5**	**-0.27**	**0.002**	–	–	IN-from-FR
	Italian	47.19	48.74	2417.5	–	0.273	–	–	–
Type-token ratio	French	55.61	59.63	2494	–	0.051	–	–	–
	Italian	60.66	59.63	1976	–	0.364	–	–	–
Core vocabulary	French	55.36	52.43	**1454**	**-0.25**	**0.003**	–	–	IN-from-FR
	Italian	52.96	52.43	2080.5	–	0.665	–	–	–
List head coverage	French	55.99	52.43	–	–	**<0.001**	**53.50**	**0.04**	IN-from-FR
	Italian	55.01	52.43	–	–	**<0.001**	**24.35**	**0.03**	IN-from-IT

TABLE 4

Translated vs. original written English

Simplification parameter	Source language	Results (M, %)		Statistical tests				Simplification in subcorpus
		TR	OR-W	W	r	p	$\chi^2(1)$	
Lexical density	French	46.72	48.27	2366	–	0.179	–	–
	Italian	49.72	48.27	**1550**	**-0.25**	**0.004**	–	OR-W
Type-token ratio	French	58.66	61.09	2342	–	0.218	–	–
	Italian	62.59	61.09	1839.5	–	0.126	–	–
Core vocabulary	French	53.81	53.03	1828	–	0.236	–	–
	Italian	51.49	53.03	**2683**	**-0.20**	**0.02**	–	OR-W
List head coverage	French	52.46	52.14	–	–	0.516	0.42	–
	Italian	51.66	52.14	–	–	0.355	0.85	–

As for list heads, the results of the χ^2 tests confirm that interpreted speeches are significantly more repetitive than non-interpreted speeches, regardless of the source language, while translations are very similar to non-translated texts. However, the effect sizes are very small, which indicates that the high significance level could be due to the large number of observations (as each individual word is classified as either "head" or "tail"), rather than to an actual effect of mediation.

Overall, evidence of lexical simplification appears to be fairly consistent for interpretations from French, but is missing for interpretations from Italian. As far as translations are concerned, those from French pattern with original written English texts, while those from Italian display an unexpected behaviour – they are more complex than comparable originals.

5.3. Intermodal comparisons

We next consider differences between interpreted and translated English (Table 5), related to our RQ2.

Table 5
Translated vs. interpreted English

Simplification parameter	Source language	Results (M,%)		Statistical tests					Simplification in subcorpus
		TR	IN	W	r	p	$\chi^2(1)$	φ	
Lexical density	French	46.72	45.54	1705	–	0.058	–	–	–
	Italian	49.72	47.19	**1405.5**	**-0.33**	**<0.001**	–	–	IN-from-IT
Type-token ratio	French	58.66	55.61	**1654**	**-0.19**	**0.033**	–	–	IN-from-FR
	Italian	62.59	60.66	1940	–	0.106	–	–	–
Core vocabulary	French	53.81	55.63	2458	–	0.108	–	–	–
	Italian	51.49	52.96	**2897.5**	**-0.20**	**0.002**	–	–	IN-from-IT
List head coverage	French	52.46	55.99	–	–	**<0.001**	**53.11**	**0.04**	IN-from-FR
	Italian	51.66	55.01	–	–	**<0.001**	**37.95**	**0.03**	IN-from-IT

In the intermodal perspective, the results are quite clear-cut: translations are consistently more complex than interpretations. However, the evidence is not equally strong for the two source languages, as the difference reaches significance in three out of four comparisons for texts mediated from Italian (type-token ratio being the odd one out), and two out of four comparisons for texts mediated from French (where the difference is significant for type-token ratio and list head coverage). List head coverage comparisons are again characterised by very small effect sizes; higher practical significance is found for the differences that concern the remaining three simplification parameters, where effect sizes can be described as medium.

5.4. Source-specific comparisons

Table 6 reports on comparisons between texts mediated from different source languages (RQ 3).

Table 6
English mediated from French vs. Italian

Simplification parameter	Mediation mode	Results (M,%)		Statistical tests				Simplification in subcorpus
		From-FR	From-IT	W	r	p	$\chi^2(1)$	
Lexical density	Interpreting	45.54	47.19	**1639.5**	**-0.22**	**0.010**	–	IN-from-FR
	Translation	46.72	49.72	**1194**	**-0.40**	**<0.001**	–	TR-from-FR
Type-token ratio	Interpreting	55.61	60.66	**1432**	**-0.30**	**<0.001**	–	IN-from-FR
	Translation	58.66	62.59	**1520**	**-0.27**	**0.002**	–	TR-from-FR
Core vocabulary	Interpreting	55.63	52.96	**2897.5**	**-0.27**	**0.002**	–	IN-from-FR
	Translation	53.81	51.49	**3104.5**	**-0.35**	**<0.001**	–	TR-from-FR
List head coverage	Interpreting	55.99	55.01	–	–	0.060	3.53	–
	Translation	52.46	51.66	–	–	0.114	2.49	–

English texts translated/interpreted from French and Italian differ along most simplification parameters; the difference is non-significant only in the case of list head coverage. The pattern is once again consistent: texts mediated from Italian are always more complex than texts mediated from French. Effect sizes are, overall, the highest found in our study.

6. Discussion of results

Summing up the results reported in Section 5, our analysis of the EPTIC corpus has shown that texts translated and interpreted from French and Italian into English display diverging patterns of lexical simplification with respect to one another and to comparable original texts. Thus, simplification is found to be contingent on both the mediation mode and the source languages involved.

Starting from the monolingual comparable analysis, Laviosa's hypothesis of greater simplification in translated language (Section 2.1) does not seem to be supported by our results, since the texts translated from French are not significantly simpler than the original written texts. Furthermore, the texts translated from Italian are significantly *more* complex than the original written texts, using fewer common words and being more lexically dense. They also show no sign of greater repetitiveness, as evidenced by the type/token ratio and list head coverage analyses.

Stronger support for the simplification hypothesis is provided instead by monolingual comparable analyses of interpretations, which indicate that interpreted English is consistently simpler than original spoken English. However, significant differences are observed for three out of four parameters in the case of texts with French as a source language (lexical density, core vocabulary and list head coverage), and for only one parameter in the case of texts with Italian as a source language (list head coverage). In this respect, if we factor in the magnitude of differences as measured by effect sizes, in no case can list head coverage be considered a reliable indicator of simplification: effect sizes are too small for the observed differences to be "meaningful or important" (Andrew, Pedersen, *et al.* 2011: 60). Thus, evidence of simplification can be said to emerge from interpretations from French, but not Italian.

Moving on to the intermodal perspective, results of comparisons follow the expected trend whereby interpretations are simpler than translations (Section 2.3), but again the trend is stronger in one language combination than the other. Texts translated from Italian display lower core vocabulary coverage and higher lexical density than corresponding interpretations, and hence are more complex than interpretations along the same dimensions that also make them more complex than original written texts. On the other hand, the only reliable evidence in favour of translations from French being more complex than corresponding interpretations is provided by the type/token ratio measure, pointing to greater lexical variety. As is the case with monolingual comparable analyses, the small effect size associated with list head coverage comparisons advises against assigning relevance to the (significant) differences observed.

Focusing specifically on source-language influence, texts translated and interpreted from French are confirmed to be significantly simpler than the corresponding texts in the same modality mediated from Italian, based on all parameters except list head coverage. The effect sizes are the highest obtained in this study. Same-modality mediated texts derived from different source languages are thus more dissimilar from each other than they are from comparable original texts, or from texts mediated in a different modality.

Several reflections can be made concerning the implications of these results. First, in terms of the variables potentially affecting simplification in translation, register variation was already suggested by Laviosa (1998b) to play an important role,

a conclusion strongly corroborated by Kruger and van Rooy (2012). Our study confirms this point with reference to EP speeches in English, which are found not to display the same trends observed by Laviosa, and in fact to partially contradict them. Looking at a single text type and a single target language, our study further emphasizes the role played by different source languages, showing how even closely related languages like French and Italian can trigger different degrees of simplification in English target texts. In this respect, our results also tally with those obtained in text categorisation experiments on the influence of multiple source languages on the target language (Section 2.2). Register and source language variables should therefore be carefully considered when designing studies that aim to detect typical features of translated language.

Extending the perspective to interpreting, our results support the hypothesis of greater simplification in interpreted texts than in comparable original speeches, though they do so only for the English-French language combination, and in terms of different parameters compared to previous work. Kajzer-Wietrzny (2012) and Sandrelli and Bendazzoli (2005) also observed that simplification patterns vary based on the language combinations considered; however, both of them found list head coverage to be the most stable parameter for distinguishing between English interpreted and original speeches. This was not the case here. Statistically significant differences in list head coverage comparisons were associated with minimal effect sizes, leading us to only consider lexical density and core vocabulary coverage as reliable indicators of simplification. Since neither Kajzer-Wietrzny (2012) nor Sandrelli and Bendazzoli (2005) report effect sizes, direct comparison with our results is impossible.

Lastly, from the intermodal angle, we do find, like Bernardini, Ferraresi, *et al.* (2016: 82), that "interpreters simplify their input more than translators do" (Hu 2016: 203-204). These differences, however, are stronger in mediation from Italian than French. Furthermore, our results are only partially consistent with the suggestion that "mediated texts are simpler than non-mediated ones" (Bernardini, Ferraresi, *et al.* 2016: 82), since this only applies to interpretations here, and in targets from French more than in targets from Italian. The picture that emerges is thus more composite than the one in Shlesinger and Ordan (2012: 54), and Bernardini, Ferraresi, *et al.* (2016: 82), according to which interpreting can be viewed as "an extreme case of translation, one in which those features that have been found to distinguish between translated and original texts […] are found to be all the more salient."

Claims about simplification in an intermodal perspective can hardly disregard source language effects, alongside those of mediation mode, hence the importance of multi-source intermodal corpora like EPTIC. Which takes us to our conclusion, and to a few final caveats.

7. Conclusion

In this article we have presented a study of lexical simplification in different modes of language mediation, from two different source languages. Relying on a new and enlarged version of EPTIC, an intermodal corpus featuring plenary speeches at the European Parliament, we compared English interpretations and translations from French and Italian to each other and to corresponding original speeches and verbatim reports.

Our observations – that simplification is contingent on both the mediation mode and the source languages involved, and that the influence of the latter might in fact be stronger than that of the former – advise against sweeping generalisations and in favour of carefully designed studies of small, homogeneous, well-documented, shareable corpora representing different variables relevant to translation.

In his discussion of the nature of explanations in translation studies, Toury (2004: 24; original emphasis) noted the "enormous *heterogeneity*" of translation practices, and that "there seems to be no single factor which cannot be enhanced, mitigated, maybe even offset by the presence of another" (Toury 2004: 15). Taking the effect of source language as a case in point, one may legitimately wonder whether this is an adequate variable in explaining the observed simplification patterns. While factoring in the complexity of source texts in different languages might take us one step further in shedding light on this point, it would not exclude other related variables, ranging from socio-cultural ones, such as the norm-setting traditions in training institutions (Toury 1995), to individual ones, such as language competence or translation and interpreting skills. We leave it to further studies to unravel this "tangled knot" and put "its different constituents […] in some *hierarchical*[13] order: more and less potent, more and less translation-specific, and the like" (Toury 2004: 25).

Despite these limits, and others such as the small size of the EPTIC corpus and the very peculiar nature of the register and of interpreting/translation practices at the European Union, we see this study as a contribution to establishing that "system of interconnected, mutually conditioning statements" that Toury (2004: 25) suggested could be "a reasonable ultimate goal for Translation Studies."

ACKNOWLEDGEMENTS

We would like to thank the issue editors for their assistance and support and the two anonymous reviewers for their extensive, insightful comments on the manuscript. But first and foremost, our thanks go to the students who took part in the construction of EPTIC: Marie De Clerck, Émilie Degueldre, Sophie Steil, Fiona Thewissen, Florent Thirion, and Coraline Zizi at the Louvain School of Translation and Interpreting (UCLouvain, Belgium); Tiziana Calà, Rita Micchi, Manuela Santandrea, and Erika Stragapede at the Department of Interpreting and Translation (University of Bologna, Italy). We also thank Dr. Nicoletta Spinolo for her help with the abstract and keywords in Spanish.

NOTES

1. Throughout this article we use the term *simplification* due to its wide currency in the field, even though *simplicity* would be more appropriate for our monolingual comparable perspective.
2. We focus on lexical measures only, and exclude average sentence length, since in previous work it has been found to vary erratically (Section 2).
3. According to several EU officials consulted on the translation practice at the EP, translations are carried out independently of the interpretations: they are entirely based on the verbatim reports, with no reference to the original oral speeches or the interpreters' renditions.
4. Plenary sessions. *EPTV*. Bruxelles: European Parliament. Visited 23 March 2018, <http://www.europarl.europa.eu/ep-live/en/plenary>.
5. Other speeches in the languages currently covered, as well as speeches in additional languages (Slovene, Polish and Finnish), are in the process of being added to EPTIC.
6. As the practice of translating verbatim reports was discontinued in the second half of 2011, speeches were selected among the most recent for which translations are available.

7. Schmid, Helmut (1995): *TreeTagger*. Visited 1 July 2018, <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>.
8. *European Parliament Translation and Interpreting Corpus*. Visited 1 July 2018, <http://corpora.dipintra.it/eptic>.
9. The extraction of running words is performed via the *cwb-scan-corpus* tool, using the regular expression [a-zA-Z0-9àèéìòùÈ\'\-]+[^\-] (where [^\-] excludes truncated words).
10. Adverbs are actually all annotated with the same tag, so we excluded from the count of function words those that end in "-ly" (except "only"). Auxiliary verbs, on the other hand, have a designated tag and were thus easily classified as function words.
11. The EPTIC tagset can be found here: <http://docs.sslmit.unibo.it/doku.php?id=corpora:tagsets:english>.
12. R Core Team (2018): *R*. Version 3.4.4. Visited 1 July 2018, <http://www.r-project.org>.
13. Emphasis in the original.

REFERENCES

Andrew, Damon P. S., Pedersen, Paul M., and McEvoy, Chad D. (2011). *Research Methods and Design in Sport Management*. Champaign: Human Kinetics.

Baker, Mona (1993): Corpus linguistics and translation studies: Implications and applications. *In*: Mona Baker, Gill Francis, and Elena Tognini-Bonelli, eds. *Text and Technology: In Honour of John Sinclair*. Amsterdam/Philadelphia: John Benjamins, 233-250.

Baroni, Marco, Bernardini, Silvia, Ferraresi, Adriano, *et al*. (2009): The WaCky Wide Web: A collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation*. 43(3):209-226.

Bernardini, Silvia, Ferraresi, Adriano, and Miličević, Maja (2016): From EPIC to EPTIC – Exploring simplification in interpreting and translation from an intermodal perspective. *Target*. 28(1):61-86.

Biber, Douglas and Jones, James K. (2009): Quantitative methods in corpus linguistics. *In*: Anke Lüdeling and Merja Kytö, eds. *Corpus Linguistics: An International Handbook*. Berlin: de Gruyter, 1286-1304.

Cappelle, Bert (2012): English is less rich in manner-of-motion verbs when translated from French. *Across Languages and Cultures*. 13(2):173-195.

Cappelle, Bert and Loock, Rudy (2017): Typological differences shining through. The case of phrasal verbs in translated English. *In*: Gert De Sutter, Marie-Aude Lefer, and Isabelle Delaere, eds. *Empirical Translation Studies: New Methodological and Theoretical Traditions*. Berlin: de Gruyter, 235-264.

Chesterman, Andrew (2004): Beyond the particular. *In*: Anna Mauranen and Pekka Kujamäki, eds. *Translation Universals. Do they Exist?* Amsterdam/Philadelphia: John Benjamins, 33-49.

Corpas Pastor, Gloria, Mitkov, Ruslan, Naveed, Afzal, *et al*. (2008): Translation universals: do they exist? A corpus-based NLP study of convergence and simplification. *In*: *MT at work: Proceedings of the Eighth Conference of the Association for Machine Translation in the Americas*. (AMTA2008: The Eighth Conference of the Association for Machine Translation in the Americas, Waikiki, 21-25 October 2008). Stroudsburg: Association for Machine Translation in the Americas, 75-81.

Defrancq, Bart, Plevoets, Koen, and Magnifico, Cédric (2015): Corpus pragmatics in translation and contrastive studies. *Yearbook of Corpus Linguistics and Pragmatics*. 3:195-222.

Ferraresi, Adriano and Bernardini, Silvia (2019): Building EPTIC: A many-sided, multipurpose corpus of EU Parliament proceedings. *In*: Irene Doval and Maite Sánchez Nieto, eds. *Parallel Corpora for Contrastive and Translation Studies: New Resources and Applications*. Amsterdam/Philadelphia: John Benjamins, 123-139.

Field, Andy, Miles, Jeremy, and Field, Zoe (2012): *Discovering Statistics Using R*. London: Sage.

HAREIDE, Lidun (2017): Is there gravitational pull in translation? *In*: Meng JI, Michael OAKES, Defeng LI, *et al*., eds. *Corpus Methodologies Explained: An Empirical Approach to Translation Studies*. London/New York: Routledge, 188-231.

HU, Kaibao (2016): *Introducing corpus-based translation studies*. Heidelberg: Springer.

HU, Kaibao and TAO, Quing (2013): The Chinese-English conference interpreting corpus: Uses and limitations. *Meta*. 58(3):626-642.

ILISEI, Iustina, INKPEN, Diana, CORPAS PASTOR, Gloria, *et al*. (2010): Identification of translationese: A machine learning approach. *In*: Alexander GELBUKH, ed. *Computational Linguistics and Intelligent Text Processing*. Heidelberg: Springer, 503-511.

KAJZER-WIETRZNY, Marta (2012): *Interpreting Universals and Interpreting Style*. Doctoral dissertation, unpublished. Poznan: Adam Mickiewicz University.

KOLEHMAINEN, Leena and RIIONHEIMO, Helka (2016): Literary translation as language contact. *Literary Linguistics*. 5(3):1-32.

KOPPEL, Moshe and ORDAN, Noam (2011): Translationese and its dialects. *In*: Yuji MATSUMOTO and Rada MIHALCEA, eds. *HLT '11: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*. (ACL 2011: 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Portland, 19-24 June 2011). Vol. 1. Stroudsburg: Association for Computational Linguistics, 1318-1326.

KRANICH, Svenja (2014): Translations as a locus of language contact. *In*: Juliane HOUSE, ed. *Translation: A Multidisciplinary Approach*. Basingstoke: Palgrave Macmillan, 96-115.

KRUGER, Haidee and VAN ROOY, Bertus (2012): Register and the features of translated language. *Across Languages and Cultures*. 13(1):33-65.

KRUGER, Haidee and VAN ROOY, Bertus (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, Sara, ed. (1998a): Special issue - The corpus-based approach: A new paradigm in translation studies. *Meta*. 43(4).

LAVIOSA, Sara (1998b): Core patterns of lexical use in a comparable corpus of English narrative prose. *Meta*. 43(4):557-570.

LAVIOSA, Sara (1998c): The English comparable corpus. A resource and a methodology. *In:* Lynne BOWKER, Michael CRONIN, Dorothy KENNY, *et al*., eds. *Unity in Diversity? Current Trends in Translation Studies*. Manchester: St. Jerome, 101-112.

LAVIOSA, Sara (2002): *Corpus-Based Translation Studies*. Amsterdam: Rodopi.

RUSSO, Mariachiara, BENDAZZOLI, Claudio, SANDRELLI, Annalisa, and SPINOLO, Nicoletta (2012): The European parliament interpreting corpus (EPIC): Implementation and developments. *In*: Francesco STRANIERO SERGIO and Caterina FALBO, eds. *Breaking Ground in Corpus-based Interpreting Studies*. Bern: Peter Lang, 53-90.

RYCHLÝ, Pavel (2007): Manatee/Bonito – A modular corpus manager. *In*: Petr SOJKA and Aleš HORÁK, eds. *Proceedings of the 1st Workshop on Recent Advances in Slavonic Natural Language Processing*. (RASLAN 2007 - First Workshop on Recent Advances in Slavonic Natural Language Processing, Brno, 14-16 December 2007). Brno: Masaryk University, 65-70.

SANDRELLI, Annalisa and BENDAZZOLI, Claudio (2005): Lexical patterns in simultaneous interpreting. *In*: Pernilla DANIELSSON and Martijn WAGENMAKERS, eds. *Proceedings from the Corpus Linguistics Conference Series*. (CL'05: Corpus Linguistics 2005, Birmingham, July 14-17 2005). Birmingham: University of Birmingham. Visited 2 February 2018, <https://www.birmingham.ac.uk/research/activity/corpus/publications/conference-archives/2005-conf-e-journal.aspx>.

SHLESINGER, Miriam (1998): Corpus-based interpreting studies as an offshoot of corpus-based translation studies. *Meta*. 43(4):486-493.

SHLESINGER, Miriam (2009): Towards a definition of interpretese: An intermodal, corpus-based study. *In*: Gyde HANSEN, Andrew CHESTERMAN, and Heidrun GERZYMISCH-ARBOGAST, eds. *Efforts and Models in Interpreting and Translation Research: A Tribute to Daniel Gile*. Amsterdam/Philadelphia: John Benjamins, 237-253.

Shlesinger, Miriam and Ordan, Noam (2012): More spoken or more translated? Exploring a known unknown of simultaneous interpreting. *Target*. 24(1):43-60.

Stubbs, Micheal (1996): *Text and Corpus Analysis*. Oxford: Blackwell.

Szmrecsanyi, Benedikt and Kortmann, Bernd (2009): Between simplification and complexification: non-standard varieties of English around the world. *In:* Geoffrey Sampson, David Gill, and Peter Trudgill, eds. *Language Complexity as an Evolving Variable*: Oxford: Oxford University Press, 64-79.

Toury, Gideon (1995): *Descriptive Translation Studies – and Beyond*. Amsterdam/Philadelphia: John Benjamins.

Toury, Gideon (2004): Probabilistic explanations in translation studies. *In*: Anna Mauranen and Pekka Kujamäki, eds. *Translation Universals. Do they Exist?* Amsterdam/Philadelphia: John Benjamins, 15-32.

van Halteren, Hans (2008): Source Language Markers in EUROPARL Translations. *In*: Donia Scott and Hans Uszkoreit, eds. *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*. (COLING 2008: The 22nd International Conference on Computational Linguistics, Manchester, 18-22 August 2008). Manchester: The Coling 2008 Organizing Committee, 937-944.

Williams, Donna (2005): *Recurrent Features of Translation in Canada*. Doctoral dissertation, unpublished. Ottawa: University of Ottawa.

Xiao, Richard and Hu, Xianyao (2015): *Corpus-Based Studies of Translational Chinese in English-Chinese translation*. Heidelberg: Springer.