

EN-terpretese: a corpus-based exploratory study¹

CRISTINA SCARDULLA
MARK DAVID MATHIAS

ABSTRACT

The present study is a corpus-based exploratory study aimed at comparing and contrasting three subcorpora of English interpreted vs. non-interpreted texts. Its general objective is to isolate any linguistic features of interpreted discourse – as opposed to original production – in the restricted set of texts analysed, which could be ascribed to *interpretese*, namely the way in which interpreters speak, irrespective of the language pair they are working in and the source text interferences they might be exposed to. The three subcorpora contain speeches delivered by native speakers in national and international settings, as well as speeches interpreted into English by professional interpreters working at the European Parliament. In the first part, the study focuses on the existing notions of *simplification* and *explicitation* as interpreting universals, to test whether interpreted speeches are actually simpler and more explicit than original ones. The second part analyses the distribution and use of different parts of speech (POS) in the three subcorpora as well as the use of figures of speech by the speakers in the three identified groups and settings. The results obtained seem to suggest that the time pressure interpreters

¹ This paper stems from the ideas of both authors. The Introduction and Sections 1, 3, 3.2, 3.3, 4.1 and 4.3 were written by Cristina Scardulla, whereas Sections 3.1, 4.2 and 4.4 by Mark David Mathias. The remaining parts (Section 2 and Conclusions) were written jointly.

are exposed to as well as the interpreting strategies adopted might play a decisive role in the production of the interpreted text. More specifically, interpreted texts, despite showing a lower degree of language variety, especially as concerns lexical words, appear to be more informative. Furthermore, their level of explicitness is similar to that of non-interpreted texts. Finally, the use of figures of speech appears to depend more on the speaker's personal style or type of event rather than on the audience it is addressed to, i.e. national or international.

INTRODUCTION

Corpus-based translation and corpus-based interpreting studies (CTS and CIS) have traditionally followed two paths, namely either comparing translated/interpreted texts to their sources, mostly focusing on translators/interpreters' strategies (interlingual approach) or analysing translated/interpreted texts in parallel with comparable original written or oral texts in the same language (monolingual approach). A third path has recently emerged (intermodal approach), first introduced by Shlesinger (1998), and aimed at studying the differences and similarities between translated and interpreted text. However, recent intermodal comparisons between translation and interpretation seem to indicate that "these two forms of mediated output are more dissimilar to each other than they are to comparable non-mediated texts" (Ferraresi / Milicevic 2017) and that interpretese would therefore be more typically spoken than translated, exhibiting "far more similarities to original speech than to written translation" (Shlesinger / Ordan 2012: 47).

Therefore, the present study only focuses on the oral dimension and sets out to identify similarities (and differences) between original English speeches and interpreted English speeches by following both a *corpus-based* and a *corpus-driven* approach, using the terms introduced by Tognini-Bonelli² in 2001. According to the author's distinction, corpus-based studies use the data in the corpus to test a hypothesis formulated in advance of the analysis (be it in the literature or by the researcher). Corpus-driven studies, on the other hand, regard the corpus itself as the only source for formulating any language hypothesis. Refuting the binary distinction between the two notions (see McEnergy / Hardie 2011: 151-152), the study adopts the term 'corpus-based' (which can also be found in the title) to encompass both labels (see

² "The term *corpus-based* is used to refer to a methodology that avails itself of the corpus mainly to expound, test or exemplify theories and descriptions that were formulated before large corpora become available to inform language study [...] corpus-based linguistics adopt a 'confident' stand with respect to the relationship between theory and data in that they bring with them models of language and descriptions which they believe to be fundamentally adequate, they perceive and analyse the corpus through these categories and sieve data accordingly [...]" (2001: 65-6); on the other hand "in a corpus-driven approach the commitment of the linguists to the integrity of the data as a whole, and descriptions aim to be comprehensive with respect to corpus evidence. The corpus, therefore, is seen as more than a repository of examples to back pre-existing theories or a probabilistic extension to an already well-defined system" (Tognini-Bonelli 2001: 84).

McEnery *et al.* 2005: 11). Hence, it integrates both approaches to explore the applicability of existing theories on interpretese in the corpus and to analyse any patterns or trends identifiable in the three subcorpora by following a bottom-up approach.

1. INTERPRETESE

It has often been observed that the product of translation is somehow different than non-translated text. Scholars have emphasised two distinct bases for such differences: some have focused on the interference from the source language spilling over into the translation on different levels (see, among others, Toury's law of interference: "in translation, phenomena pertaining to the make-up of the source text tend to be transferred to the target text", 1995: 275), whereas others have noted general effects of translation process, irrespective of the source language, and have searched for unique and inherent patterns of linguistic behaviour in translation, commonly known as translation universals (see, among others, Baker 1996, Olohan 2002, Toury 2004, Chesterman 2004, Zanettin 2012). Corpus linguistics has proven to be the most effective methodology to identify these features, "which typically occur in translated texts rather than original utterances and which are not the result of interference from specific linguistic systems" (Baker 1993: 243). Machine-readable corpora have made it possible to formulate general assumptions about translated language. One of the first corpus-based studies focused on simplification as a translation universal was carried out by Laviosa (1998), who built a monolingual comparable corpus of original and translated English texts. The author discovered that the variety of vocabulary used in the corpus of translated texts was more limited, irrespective of the source language, thus confirming the relevance of simplification as a distinctive feature of translation.

Corpora found their way into interpreting studies slightly later, mainly due to the difficulties related to the initial limited availability of real-life material and the technical hurdles of the transcription phase. In 2005, Sandrelli and Bendazzoli replicated Laviosa's methodology to test the same simplification hypothesis on a set of subcorpora of EPIC (European Parliament Interpreting Corpus), more specifically the subcorpus of interpreted English texts, the one of interpreted Italian texts and two subcorpora of original speeches in those two languages. Even though the authors' results only partially confirm Laviosa's observations, it has to be noted that their reference corpora also contained texts by non-native speakers of English (both in the original speakers' and the interpreters' groups) and that the corpora were very uneven in size, which might have somewhat influenced the results. A more recent and larger study by Kajzer-Wietrzny (2012) on 'interpreting universals' set out to test the features of simplification, explicitness and normalisation applying translation-studies parameters to a corpus including both interpretation and translation into English from four different languages (French, Spanish, German and Dutch) and speeches originally produced in English. The three key

parameters selected for testing the simplification hypothesis were: list heads, lexical density and the proportion of high frequency words. The analysis revealed that interpreting was in fact less lexically simplified than original English speeches, though the author points out that “the mode of delivery of the source speech might have an influence on the level of lexical sophistication of the target text” (Kajzer-Wietrzny 2012: 98). Furthermore, the study highlighted considerable differences across interpreted speeches from different languages, making it difficult to formulate any generalisation on the principle of simplification.

As to the principle of explicitation, much research has been carried out in the past on the use of the optional complementiser *that*, both in general in the English language and more specifically by professional translators and interpreters. A seminal study by Olohan and Baker (2000) found that translators tend to use the optional complementiser *that* more often than authors in non-translated texts, in other words translators tend to be more explicit than speakers who produce an original, non-translated text. Similarly, Kajzer-Wietrzny (2012) has shown that interpreters tend to be more explicit than native English speakers, as they resort to the optional complementiser *that* more often than native English MEP speakers.

There are many reasons why a speaker may resort to the optional complementiser *that* in their speech or may instead opt for the zero complementiser: formality and informality, higher or lower frequency of the matrix verb, and the complexity of the subordinate clause are three major categories (Olohan / Baker 2000: 150). Many authors have argued and demonstrated that the more common the matrix verb is, the less the complementiser *that* is used, whereas the less common the matrix verb is, the more the complementiser *that* is used. Indeed, Rohdenburg (1996: 151) explains this pattern by arguing that a more complex matrix verb entails a greater cognitive burden for the speaker, who will consequently feel the need to add the optional *that* to their sentence. Finally, the third main reason why an optional *that* may be used is the presence of a complex sentence structure: Elness (1984: 519) argues that a high number of words present in the object clause will most likely lead to the speaker connecting the main clause and the object clause with an optional complementiser *that*. One of the goals of the present study is to determine whether there were remarkable quantitative differences in the use of the optional complementiser *that* between the three subcorpora and to test Kajzer-Wietrzny's conclusions on this particular corpus.

2. METHODOLOGY

The study follows a two-pronged approach in order to both test existing theories on interpreting in the corpus and analyse any patterns directly identifiable in the three subcorpora (see Introduction). More specifically, the study focuses on the existing notions of *simplification* (“the idea that translators subconsciously simplify the language or message or both”, Baker 1996: 176) and *explicitation* (“an

overall tendency to spell things out rather than leave them implicit in translation”, Baker 1996: 180-181) respectively, which have been put forward as translation universals.

Furthermore, a selection of features are analysed, which resulted from a bottom-up approach to the corpus, in that the data from two different areas of linguistic analysis have been collected – namely part-of-speech (POS) distribution (see 4.3) and use of figures of speech (see 4.4), to verify if any salient traits would emerge, which could then be tested in future studies using a more ‘corpus-based’ approach (in the original meaning of Tognini-Bonelli 2001; see Introduction).

In order to conduct the quantitative analysis, the corpus manager and text analysis software Sketch Engine has been used, which also offers automatic POS-tagging by means of the TreeTagger tool.

The indicators selected to test simplification are Type-Token Ratio (TTR), high frequency words and lexical density and have been calculated using Sketch-Engine generated data.

For the indicator of explicitation, the use of the optional complementiser *that* was studied in the three subcorpora. Sketch Engine allowed for the identification of all occurrences of *that*. Subsequently, the occurrences were selected, divided into separate categories and compared (see 4.2).

As to the analysis on the distribution of different part of speech, both data stemming from Sketch Engine and data resulting from a more in-depth study of the texts have been used. A first distinction between lexical and functional words has been made by resorting to the Sketch Engine’s automatic POS-tagging tool, which was then complemented by a qualitative analysis of the categories of nouns, verbs, adverbs and adjectives in the texts (see 4.3).

As to the figures of speech, data obtained from in-depth study of the texts have been used. Three figures of speech (the anaphora, the tricolon and the antithesis) were identified in the texts and subsequently compared (see 4.4).

3. THE *EN-terpretese* CORPUS

The corpus used for the present study is constituted of three comparable subcorpora. Comparable corpora consist of a collection of texts in one language and translations in the very same language from one or more source languages. More specifically the *EN-terpretese* corpus consists of three subcorpora:

1. The *EN national subcorpus*: a collection of speeches delivered in English by native speakers and addressed to a national audience;
2. The *EN European subcorpus*: a collection of speeches delivered in English by native speakers and addressed to a European audience; and
3. The *EN interpreted subcorpus*: a collection of speeches interpreted from Italian into English by native speakers and addressed to a European audience.

All subcorpora are composed of around 20,000 words. The speeches building the corpus have been selected to respond to a series of criteria, which would guarantee texts comparability. First of all, they are all delivered in comparable settings, mostly in Parliaments (more specifically the House of Commons and the European Parliament) and by actors enjoying the same roles and status³ (e.g. national or European Members of Parliament, European Commissioners, Prime Ministers). These two factors entail that the audience that the speakers are addressing is also comparable (e.g. fellow members of Parliament, legislators, and the greater audience of voters, as all selected events were web-streamed or recorded to be later published online). Furthermore, all subcorpora include both more formal, prepared remarks and spontaneous speech, to account for all delivery modes which typically feature in the selected communicative events. As the present study does not focus mainly on lexical choices and does not include a terminological analysis, no priority has been assigned to the selection of texts by topic. Single texts vary in length and even though most of the topics dealt with are recurrent, there is not a perfect correspondence of themes in all the subcorpora.

Comparing oral texts as opposed to comparing written texts presents a series of specific hurdles, mostly in terms of the transcription process and the “near impossibility of incorporating the full gamut of paralinguistic and prosodic features” (Shlesinger 2009: 239). As this is an exploratory study focusing on language patterns, rather than suprasegmental and prosodic features, pauses and hesitations, all these elements have not been accounted for in the transcription.

A. THE EN NATIONAL SUBCORPUS

The EN national subcorpus consists of 20,199 words. The speeches are delivered by 14 speakers, including one Prime Minister, a Secretary of State for Environment, Food and Rural Affairs, a Shadow Secretary of State for Environment and Rural Affairs and several MPs. The speeches are taken from four events: the Mansion House Speech, a Liaison Committee meeting, the 2018 Conservative Party Conference and a Backbench Business Committee debate. Speeches either take the form of formal and prepared statements or of spontaneous speech. The topics covered in the speeches are all country-specific: the achievements of the British Conservative Party, Brexit, its effects on British agriculture, the UK’s potential future trade deals with other countries and the UK’s post-Brexit strategy.

Table 1 offers a detailed overview of the EN interpreted subcorpus.

³ Professional interpreters, by definition, always enjoy the same role and status, namely that of interpreter, but they are trained to adjust their delivery and output to match the role and status of the speaker they are interpreting.

Speech Code	Speaker	Event	Text type	Topic	Word count
N__S1.1 ⁴	UK Prime Minister	Mansion House speech	Speech by the UK Prime Minister on the UK's future economic partnership with the EU	Brexit, future EU-UK relations, rulings of the European Court of Justice.	6,815
N__S1.2	UK Prime Minister	Liaison Committee ⁵ meeting	Statement to the committee	Brexit, withdrawal agreement	695
N__S2	MP		Question to the Prime Minister	No-deal Brexit scenario	77
N__S1.2	UK Prime Minister		Reply to the MP	Commitment of UK government to reach a deal	169
N__S3	Secretary of State for Environment, Food and Rural Affairs	2018 Conservative Party Conference	Speech at Conservative Party Conference	Environment, DEFRA, future after Brexit, achievements of Conservative Party	2,272
N__S4	MP	Backbench Business Committee ⁶ debate: "The policy framework for agriculture after the UK leaves the EU".	Remarks on CAP, Brexit	CAP, Brexit, agriculture, food prices	2,393
N__S5	MP		Remarks on Brexit effects on the UK agriculture	Environmental protection and post-Brexit agricultural policy	743
N__S6	MP		Remarks on post-Brexit strategy for the UK	Subsidies and agricultural jobs	854
N__S7	MP		Remarks on post-Brexit farming	Brexit, farming, food production	1,532
N__S8	MP		Remarks on UK agriculture	Agriculture and environmental protection	889
N__S9	MP		Remarks on agriculture and related issues	UK framework for agriculture, funding, the environment and infrastructure	870
N__S10	MP		Remarks on UK framework for agriculture	Agriculture and labour in the agricultural sector	1,187
N__S11	MP		Remarks on Government role in agriculture and food production industries	Agriculture and food production industries	722
N__S12	Shadow Secretary of State for Environment and Rural affairs		Remarks on importance of agriculture for the UK	Agriculture and UK trade agreements	981
					20,199

Table 1: Overview of the sub-corpus of texts by English native speakers addressing a national audience.

4 In the code N__S, N stands for 'national' and S for 'speech', followed by a unique code. A second digit is added when the corpus contains more than one speech by the same speaker.

5 "The Liaison Committee is made up of Select Committee Chairs. It considers the overall work of select committees, promotes effective scrutiny of Government and chooses committee reports for debates. It questions the Prime Minister about policy, usually three times a year" (see <<https://committees.parliament.uk/committee/103/liaison-committee-commons>>, last accessed August 2021).

6 "The Backbench Business Committee gives opportunities to backbench Members of Parliament to bring forward debates of their choice. It was the first business committee of any kind to be established by the House of Commons" (see <<https://committees.parliament.uk/committee/202/backbench-business-committee>>, last accessed August 2021).

B. THE EN EUROPEAN SUBCORPUS

The EN European subcorpus consists of 20,029 words. Speeches are delivered by ten native speakers, who all enjoy equal status as national or European officials and politicians of the highest ranks (Prime Ministers, European Commissioners, Members of the European Parliament). Speeches are delivered in an institutional context, that is public events, open to a multilingual and multicultural audience, comprising both direct participants and interlocutors and the broader public of European citizens, as all events are either web-streamed or recorded to be then published online. The type of events selected range from European Parliament committee meetings and plenary sessions, to press conferences and international forums. Speeches either take the form of formal and prepared statements or of spontaneous speech. The topics dealt with range from EU-related legislation (e.g. Common Agricultural Policy, cybersecurity, food sustainability and energy security) to country-specific issues, most notably Brexit. Table 2 offers a detailed overview of the EN European subcorpus.

Speech code	Speaker	Event	Text type	Topic	Word count
EUR_S1 ⁷	Ireland Prime Minister	European Parliament plenary session	Opening statement on the Future of Europe (series of future of Europe debates)	Overview on the Irelands' role in the EU	3,234
EUR_S2	UK Prime Minister	Press conference at the European Parliament	Official statement by UK Prime Minister on Brexit	Brexit, Northern Ireland backstop	674
EUR_S3	MEP	European Parliament plenary session	Intervention in plenary on the Brexit vote	Brexit, future EU-UK deal	755
EUR_S4.1	EU Commissioner for Agriculture and Rural Development	Committee on Agriculture (AGRI), extraordinary meeting	Presentation of the EU Commissioner on future legislative proposals	Common Agricultural Policy, environment and climate action	4,283
EUR_S5	MEP		Question to the Commissioner		358
EUR_S4.2	EU Commissioner for Agriculture and Rural Development		Reply to the MEPs		2,281
EUR_S6	EU Commissioner for the Security Union	Disinfo Week ⁸	Closing remarks	Transatlantic relations, disinformation, cybersecurity	1,869
EUR_S4.3	EU Commissioner for Agriculture and Rural Development	International Forum on Food and Nutrition ⁹	Official remarks in an EU-related panel	Food sustainability and Common Agricultural Policy	1,776
EUR_S7	MEP	Committee on Petitions (PETI), ordinary meeting	Presentation of a petition on behalf of the petitioner	Environmental issues.	1,742
EUR_S8	MEP	Committee on Constitutional Affairs (AFCO), ordinary meeting	Comment on the vote of the Committee on the EP recommendation on Britain's Withdrawal Agreement	Brexit referendum and Brexit deal	640
EUR_S9	MEP (rapporteur)	Press conference at the European Parliament	Statement by the rapporteur on language equality in the digital age	Languages and digital technology	771
EUR_S10	MEP	Event launching the EU Energy Poverty Observatory ¹⁰	Statement in panel of EU policymakers	Energy poverty, security and efficiency	1,646
					20,029

Table 2: Overview of the sub-corpus of speeches by English native speakers addressing a European audience.

7 In the code EUR_S, EUR stands for 'European' and S for 'speech', followed by a unique code. A second digit is added when the corpus contains more than one speech by the same speaker.

8 Event organised in Brussels by the Atlantic Council, which promotes constructive leadership and engagement in international affairs based on the Atlantic Community's central role in meeting global challenges. (see <<https://www.atlanticcouncil.org/about/>>, last accessed June 2020).

9 Event organised by the Barilla Centre for Food and Nutrition (BCFN) Foundation, which is a multidisciplinary, independent, non-profit organization (see <<https://www.barillacfn.com/en/>>, last accessed June 2020).

10 The EU Energy Poverty Observatory (EPOV) was established by the European Commission and is managed by a consortium of organisations.

C. THE EN INTERPRETED SUBCORPUS

The EN interpreted subcorpus consists of 19,732 words. The original speeches are all delivered in Italian by 15 speakers, who all enjoy equal status as national or European officials and politicians of the highest ranks (Prime Ministers and Members of the European Parliament). Speeches are all delivered at the European Parliament and then interpreted into English by professional interpreters working for the EP. The type of events selected – which were all web-streamed – include European Parliament committee meetings, plenary sessions, press conferences and group meetings. Speeches either take the form of formal and prepared statements or of spontaneous speech. The topics dealt with range from EU-related legislation (e.g. unfair trade practices in the agri-food chain, tax policies, railway transport) to country-specific issues, most notably concerning Italy. Table 3 offers a detailed overview of the EN interpreted subcorpus.

Speech code	Original speaker	Event	Text type	Topic	Word count
INT__S1.1 ¹¹	Italian Prime Minister	European Parliament plenary session	Opening statement on the future of Europe (series of future of Europe debates)	Overview on the main reforms undertaken by Italy and comments on topical issues	4,419
INT__S2	EP President		Reply to PM		424
INT__S3	MEP		Reply to PM		610
INT__S4	MEP		Reply to PM		880
INT__S5	MEP		Reply to PM		676
INT__S6	MEP		Reply to PM		458
INT__S7	MEP		Reply to PM		223
INT__S1.2	Italian Prime Minister		Reply to PM		2,285
INT__S8	MEP (rapporteur)	Press conference	Presentation by the rapporteur after the EP approval of the shareholders' rights directive	Tax policy and corporate social responsibility	2,197
INT__S9	Journalist		Question to the MEP		139
INT__S10.1	MEP (rapporteur)	Press conference	Update on recent trilogue	Unfair trade practices (agri-food)	1250
INT__S11	Journalist		Question to the MEP		145
INT__S10.2	MEP (rapporteur)		Reply to journalist		664
INT__S12.1	MEP (rapporteur)	Press conference	Presentation by the rapporteur after the vote on the recast of the first railway package	Railway transport	1,439
INT__S13	Journalist		Question to the MEP		83
INT__S12.2	MEP (rapporteur)		Reply to journalist		535
INT__S14	MEP (SeD President)	Group briefing	Briefing on the SeD bureau meeting	Update to the SeD group on the last bureau meeting	981
INT__S15	MEP (rapporteur)	Press conference	Presentation by the rapporteur after vote in plenary on the packages on circular economy	Circular economy	1,169
INT__S10.3	MEP (rapporteur)	Press conference	Update on recent trilogue	Unfair trade practices (agri-food)	1,155
					19,732

Table 3: Overview of the subcorpus of speeches by interpreters working from Italian into English.

¹¹ In the code INT__S, INT stands for 'interpreted' and S for 'speech', followed by a unique code. A second digit is added when the corpus contains more than one speech delivered by the same speaker.

4. RESULTS

In this section results are presented. More specifically, section 4.1 and 4.2 focus on the existing notions of *simplification* and *explicitation* respectively (see Section 1). In the subsequent sections, the results of the analysis of part-of-speech (POS) distribution (see 4.3) and use of figures of speech (see 4.4) are presented.

4.1 LEXICAL VARIETY, REPETITIONS AND LEXICAL DENSITY AS INDICATORS OF SIMPLIFICATION

Three indicators have been selected to test simplification, namely lexical variety (Type-Token Ratio, TTR), high frequency words and lexical density.

Lexical variety is a parameter measuring how rich the vocabulary of a text is, thus representing an indicator of linguistic complexity. It is calculated by the ratio of types (non-repeated words in a subcorpus) to tokens (the total subcorpus word count). The final result is then multiplied by 100, in order to express it as a percentage. The higher the ratio, the more varied the language. The TTR range falls between a theoretical 0 (infinite repetition of a single type) and 100% (not a single word is ever repeated), which entails that a lower TTR in interpreted texts would correspond to a lower lexical variety, which in turn could be the result of a simplification process (see Zanettin 2012: 14-15).

As shown in Table 4, the three subcorpora present very similar data in terms of lexical variety: values range from 15.42% (EN national subcorpus) to a maximum of 16.46% (EN European subcorpus), with the EN interpreter subcorpus in the middle with a score of 15.76%. This indicator alone is therefore not sufficient to either confirm or refute the relevance of the simplification universal, as values only slightly vary in between subcorpora.

	EN NAT	EN EUR	EN INT
TOKENS	20,199	20,029	19,732
TYPES	3,116	3,298	3,110
TTR	15.42 %	16.46%	15.76%

Table 4: TTR and lexical variety.

The second indicator selected to test simplification is high frequency words, so as to verify, as suggested by Laviosa (1998) for translated texts, whether interpreted texts feature a higher proportion of high-frequency words. To this end, the 100 most frequent words (and respective occurrences) in the frequency lists of the three subcorpora have been identified to calculate what percentage they represent in each subcorpus. Unlike the TTR, in the case of high frequency words it is a higher value which indicates a more repetitive language and therefore lexical simplification. Additionally, each list head is further broken down into lexical

words and function words,¹² to identify which category the three subcorpora's most frequent words belong to. Data are presented in Table 5.

	EN NAT	EN EUR	EN INT
LIST HEAD WORD COUNT (100 words)	11,741	11,342	11,412
% OF SUBCORPUS	58.12%	56.62%	57.83%
LEXICAL WORDS IN LIST HEAD			
word count	3,437	3,111	3,318
% of list head	29.27%	27.42%	29.07%
FUNCTION WORDS IN LIST HEAD			
word count	8,304	8,231	8,094
% of list head	70.72%	72.57%	70.92%

Table 5: List head word count.

The data pertaining to list heads confirm the results from the lexical variety test, in that the EN national subcorpus presents the highest value and is therefore the most repetitive (the most frequent 100 words account for 58.12% of the subcorpus) and the European subcorpus the lowest value (56.62%), with the EN interpreted subcorpus in the middle with a value of 57.83%. The breakdown into lexical and function words reveals that, in addition to being the most repetitive, the head list of the EN national subcorpus also presents the highest proportion of lexical words to function words of the three. This might indicate that certain topics are probably more recurrent in the texts building the EN national subcorpus and therefore repetitions are attributable to speakers all addressing similar issues rather than to simplification. As a matter of fact, *EU* and *UK* are among the most recurrent lexical words in the subcorpus with 113 and 98 occurrences respectively, and the *EU-UK* relations have indeed dominated the national political debate across the board in recent years.

The last method used to test simplification in the corpus is that of lexical density, which is a parameter measuring the information load of a text and “is a function of the ratio of lexical (content) words to grammatical (function) words” (Stubbs

12 Function words are those words belonging to a ‘closed class’, namely those words “that have relatively fixed membership. For example, prepositions are a closed class because there is a fixed set of them in English; new prepositions are rarely coined. By contrast nouns and verbs are open classes because new nouns and verbs are continually coined or borrowed from other languages [...]” (Jurafsky / Martin 2004: 3). Lexical words, on the other hand, in addition to being part of an ‘open class’, are mostly responsible for carrying meaning. Lexical words are those falling within the categories of nouns, adjectives, verbs and adverbs.

1996: 172). The higher the percentage of lexical words in a text, the higher its lexical density and level of informativeness, whereas the lower the values, the less the text is lexically dense and therefore ‘simplified’. Results are presented in Table 6 and indicate both the number of unique items for each category and the Total Absolute Frequency (TAF), which is then used to calculate the lexical density.

		EN NAT	EN EUR	EN INT
TOKENS		20,199	20,029	19,732
NOUNS	n. items	1,207	1,280	1,181
	TAF	4,880	5,118	4,733
VERBS	n. items	475	516	497
	TAF	3,563	3,437	3,671
ADJECTIVES	n. items	448	477	447
	TAF	1,556	1,604	1,581
ADVERBS	n. items	201	205	194
	TAF	1,205	1,149	1,291
TOTAL LEXICAL WORDS		11,204	11,308	11,276
LEXICAL DENSITY		55.46%	56.45%	57.14%

Table 6: Lexical word count and lexical density.

Data show that lexical density in simultaneously interpreted texts is higher than that of speeches originally produced in English. The EN national subcorpus, in addition to being the least lexically varied and the most repetitive, is also found to be the less lexically dense, with lexical words making up 55.46% of the total word count, compared to 56.45 % of the EN European subcorpus and 57.14% of the EN interpreted subcorpus.

The tests conducted for the simplification universal do not seem to confirm its applicability to interpreted texts. A similar conclusion was reached by Sandrelli and Bendazzoli (2005), who attribute their results to the specific text-production conditions in which interpreters operate. The authors argue that “the parallel co-existence of source and target speeches and the time constraints under which interpreting is performed may explain why the patterns observed by Laviosa in relation to written texts do not apply” (Sandrelli / Bendazzoli 2005: 15).

Similarly, Kajzer-Wietrzny (2012) recorded an increased lexical density in interpreted texts, which she too attributes to interpreting-related constraints, and

more specifically to “interpreters’ avoidance of redundancy and the need to produce a very compact and dense text caused by severe time constraint in simultaneous interpreting” (2012: 95). To keep up with the speaker, interpreters might have to condensate the message and avoid repetitions, thus producing a more lexically dense speech.

When considering all three indicators, the general trend seems to indicate that there is a slight variation in lexical variety and density in simultaneously interpreted texts and speeches originally produced in English and that, unlike English translated texts (see Scarpa in this volume), interpreted discourse does not seem to confirm simplification as a universal feature of interpretese.

4.2 EXPLICITATION: THE CASE OF THE OPTIONAL THAT

This part of the study focuses on the use of what is known as the “optional complementiser *that*” in the three subcorpora. *The Cambridge International Dictionary of English* (Procter 1995: 1507) defines it as a “conjunction used to introduce a clause which reports something or gives further information, although it can often be omitted”.

The optional *that* occurrences were divided into four groups based on the following categories: formality/informality, high/low frequency of the matrix verb, simple/complex structure of the object clause and ‘other’. The final analysis entailed selecting three low frequency matrix verbs and three high frequency matrix verbs present in the corpus and comparing them with the same verbs followed by the zero complementiser present in the corpus, so as to determine whether any remarkable differences could be noted.

Table 7 shows the number of instances where the optional *that* was used in the three subcorpora:

	Optional <i>that</i>
EN NAT	146
EN EUR	125
EN INT	87

Table 7: List of optional *that* occurrences in the three subcorpora.

This part of the analysis therefore showed that, unlike in Kajzer-Wietrzny’s 2012 study, the interpreters, taken as a single group in this particular study, resorted to the optional complementiser *that* less than in both European and national subcorpora. Indeed, there were only 87 optional *that* occurrences present in the EN interpreted subcorpus, compared to 146 in the EN national subcorpus and 125 in the EN European subcorpus.

Subsequently, the single optional *that* occurrences were divided into the above-mentioned categories, namely formality and informality, lower or higher frequency of matrix verb, and the complexity of the object clause. Table 8 shows the use of the optional *that* in instances of formal sentences, presence of less frequent matrix verbs, complex subordinate clauses or instances that did not fall under any of the three previous categories, “other”.

A clause was classified as formal if a high register was employed by the speaker. In the EN interpreted subcorpus one example is:

INT (1)

Somebody said that Italy doesn't want to undertake the necessary reforms (INT__S1.2)

A clause was classified as one with a lower frequency matrix verb when the verb used was not what may be classified as a high frequency verb, e.g. *think*, *know*, *say* and *tell*, which are most frequently followed by the zero complementiser (Tagliamonte / Smith 2005: 302).

One example from the EN European subcorpus is the following:

EUR (1)

[...] and to *ensure* that into the future that parliaments and governments are more connected to each other (EUR__S1)

A clause was classified as one with a complex structure when its number of words appeared to be high in comparison to the others.

An example from the EN European subcorpus is the following:

EUR (2)

And so, I was delighted that my report was adopted by the plenary today with five nine two votes in favour, 45 against and 44 abstentions (EUR__S9)

Finally, the fourth category, named “other”, included instances in which the speaker's language was not deemed formal, no low frequency matrix verbs were used, and the object clause was neither structurally complex nor long in terms of word count.

An example of this from the EN European subcorpus is the following:

EUR (3)

And I would remind ehm I hope ultimately that Brexit, you know, well, I campaigned to remain, I believe the value for the Northeast of England, [...] (EUR__S7)

Table 8 illustrates the results for each subcorpus.

	Formality	Lower frequency of matrix verb	Complex subordinate clause	Other
EN NAT	66 (45.2%)	62 (42.5%)	8 (5.5%)	10 (6.8%)
EN EUR	60 (48%)	53 (42.4%)	11 (8.8%)	1 (0.8%)
EN INT	44 (50.6%)	34 (39.1%)	9 (10.3%)	0 (0%)

Table 8: List of optional *that* occurrences in the four main categories of use.

The decision as to whether an optional *that* was used owing to formality, to the presence of a low frequency matrix verb or to the complexity of the object clause in question was sometimes subject to a degree of subjectivity, as these can at times either coexist or overlap. The results show that, especially when it comes to the EN national and EN European subcorpora, the figures are quite similar, which arguably indicates a convergence in style between the two subcorpora. The figures in the EN interpreted subcorpus are lower; however, there is a similar distribution among the first three categories, which is proportionate to the total number of optional *that* occurrences in this subcorpus. One last remark concerns the relatively high number of clauses falling under the fourth category from the EN national subcorpus (10): these sentences were often short, entailed low register words or false starts, arguably features of speakers who feel sufficiently comfortable to use average register words, to improvise during their speeches or to change the structure of their sentences mid-sentence.

The following tables illustrate the number of cases in which three low frequency verbs (*to ensure*, *to claim* and *to recognise*) and three high frequency verbs (*to hope*, *to know* and *to say*) either were or were not followed by an optional *that*. The verbs *to claim* and *to recognise* were selected as low frequency verbs after identifying them in the subcorpora, they were not present in all subcorpora, but there were no low frequency verbs, apart from *to ensure*, that were identified in all subcorpora.

Low frequency Matrix verbs	EN interpreted		EN national		EN European	
	<i>That</i>	Z.c.	<i>That</i>	Z.c.	<i>That</i>	Z.c.
Ensure	14	2	26	7	21	2
Claim	1	0	0	0	4	0
Recognise	2	0	0	0	2	0

Table 9: List of low frequency matrix verbs followed by the optional *that* or the zero complementiser.

High frequency Matrix verbs	EN interpreted		EN national		EN European	
	<i>That</i>	Z.c.	<i>That</i>	Z.c.	<i>That</i>	Z.c.
Hope	7	4	4	1	7	1
Know	2	5	10	10	10	8
Say	10	2	9	4	6	3

Table 10: List of high frequency matrix verbs followed by the optional *that* or the zero complementiser.

The figures concerning the verb *to ensure* confirm what previous scholars have already posited, namely that the verb in question is more often followed by the optional *that* than by a zero complementiser. The interpreters may have opted for the optional *that* owing to the cognitive effort being made as they interpret, but they may have also been applying an interpreting strategy and eliminating the optional *that* as much as they could, however these hypotheses cannot be proven based on this subcorpus. The speakers from the EN national and EN European subcorpora, on the other hand, may have opted to do so owing to the low frequency of the matrix verb.

On the other hand, the figures concerning the other low frequency verbs, i.e. *to claim* and *to recognise*, are quite low, the trends noted here are the following: the speakers always opt for the optional *that* when using these verbs; the speakers from the EN national subcorpus do not use either *to claim* or *to recognise* in their speeches.

Regarding the high frequency verb *to know*, it can be noted that the speakers from the EN national subcorpus have opted half of the times for the optional *that* and half for the zero complementiser, and so did the speakers from the EN European subcorpus, despite a slightly higher number of optional *that* occurrences than zero complementisers in their case.

Furthermore, when it comes to the matrix verbs *to hope* and *to say*, the same explicitness created through the use of the optional *that* shown by the interpreters was witnessed in the other two subcorpora: in all three subcorpora the optional *that* is used more often than the zero complementiser.

4.3 PART-OF-SPEECH DISTRIBUTION

In addition to allowing for a calculation of lexical words – and therefore lexical density – the POS-tagging makes it possible to observe the lexical variety within each single identified category and the distribution of words in the corpus.

Table 11 offers an overview of the four categories identified as ‘lexical words’ for the lexical-density calculation, namely nouns, verbs, adjectives and adverbs (see Section 3.1). In addition to indicating the number of items and total absolute

frequency (TAF) for each category, data are also provided on the share of each category within the respective subcorpus' total lexical words.

	EN NAT	EN EUR	EN INT
TOT LEXICAL WORDS	11,204	11,308	11,276
% tot word count	55.46%	56.45%	57.14%
NOUNS			
n. items	1,207	1,280	1,181
TAF	4,880	5,118	4,733
% lexical words	43.55%	45.25%	41.97%
VERBS			
n. items	475	516	497
TAF	3,563	3,437	3,671
% lexical words	31.80%	30.39%	32.55%
ADJECTIVES			
n. items	448	477	447
TAF	1,556	1,604	1,581
% lexical words	13.88%	14.18%	14.02%
ADVERBS			
n. items	201	205	194
TAF	1,205	1,149	1,291
% lexical words	10.75%	10.16%	11.44%

Table 11: POS analysis of lexical words.

The relative distribution of all the categories within each single subcorpus is similar, nouns being the most frequently used part of speech in all three cases, followed in descending order by verbs, adjectives and adverbs. Nonetheless, a comparative analysis among the three subcorpora shows a more nuanced picture. The share of nouns in both the EN national and European subcorpora is higher than in the EN interpreted subcorpus (43.55%, 45.25% and 41.97% respectively). Conversely verbs account for a larger share in the EN interpreted subcorpus (32.55% against 31.80% for the EN national and 30.39% for the EN European). The tendency to resort less to nouns on the part of the interpreters is further confirmed by an analysis on the number of noun phrases of the type 'noun+noun' and 'noun+noun+noun', calculated using the Sketch Engine feature 'common query language' (CQL), as shown in Table 12.

	EN NAT	EN EUR	EN INT
NOUN+NOUN	682	832	527
NOUN+NOUN+NOUN	72	115	32
TOT	754	947	559

Table 12: Distribution of noun phrases in the three subcorpora.

The list of the 10 most recurrent lemmatised¹³ nouns (see Table 13) in the three subcorpora shows that interpreters use a similar selection of nouns as national and European speakers in terms of semantic field. Despite the fact that the topics selected for the speeches building the corpora do not perfectly overlap, the most recurrent themes seem to coincide, namely the role of the EU, its relation to single Member States (UK for national speakers and Italy for the Italian original speakers in the EN interpreted subcorpus), the role of the State, the country and the people.

	EN NAT		EN EUR		EN INT	
1	eu	132	State	93	europe	115
2	uk	107	member	92	country	79
3	food	77	farmer	87	italy	69
4	hyperlink	64	commission	60	people	52
5	country	57	europe	59	member	47
6	people	51	union	59	state	47
7	agreement	51	energy	49	parliament	46
8	policy	50	country	46	term	45
9	farmer	49	people	44	directive	41
10	market	48	eu	44	time	40
TOT		686 (14.05%)		633 (12.36%)		541 (11.43%)

Table 13: List of the 10 most recurrent lemmatised nouns in the three subcorpora.

A selection of examples on how these nouns are used shows that even though the words used are often the same, the rhetoric, the narration and the connotations differ considerably, suggesting that the audience that the speaker is addressing plays a pivotal role in the framing of the selected topics:

¹³ Sketch Engine provides lists of lemmatised nouns, verbs, adjectives and adverbs, meaning that they include all forms of the word in the results (e.g. all occurrences of go, went, gone, goes, going are counted together and listed as go).

NAT (1)

Second, the new agreement we reach with the *EU* must endure. (N__S1.1)

NAT (2)

Both the UK and the *EU* are clear: this implementation period must be time-limited. (N__S1.1)

NAT (3)

Leaving the *EU* will enable us to control the number of people entering and leaving the UK. (N__S6)

EUR (4)

This is a way that we rebuild the trust with areas which had very strong leave votes, where people have a cynicism about what the *EU* does in their daily lives. And this is a very concrete example where you can demonstrate the real value of our *EU*. (EUR__S7)

EUR (5)

That vision delivered peace in *Europe* and opened the door to peace and prosperity in my own country. (EUR__S1)

EUR (6)

A *Europe* worth building is a *Europe* worth defending. (EUR__S1)

INT (2)

We go on believing that the *Europe* that we founded is our home. (INT__S5)

INT (3)

We do not want any backsliding, but certainly we do not want a kind of *Europe* which is divided into countries moving ahead more quickly and other countries lagging behind. (INT__S14)

As to the category of verbs, even though interpreters use them more recurrently than speakers from the other two subcorpora, they only resort to a wider selection thereof than speakers in the national subcorpus (497 items in the EN interpreted subcorpus as against 475 for the national and 516 for the European, see Table 11). Table 14 shows the first ten most recurrent lemmatised verbs in the three subcorpora, thus offering a list head of the highest frequency verbs. Not only do the first ten verbs account for 50% of all verbs used in the EN interpreted subcorpus, but the first three verbs (*be*, *have* and *do*) alone account for 40.26% of the whole verb list (35.89% for the national and 39.59% for the European), thus confirming a high degree of repetitiveness in the use of verbs.

	EN NAT		EN EUR		EN INT	
1	be	847	be	893	be	941
2	have	296	have	357	have	431
3	do	136	do	111	do	106
4	need	79	make	57	think	90
5	want	76	want	41	see	54
6	make	70	get	40	go	49
7	think	64	work	40	say	48
8	go	52	take	40	make	42
9	leave	45	see	37	want	41
10	say	41	come	37	talk	40
TOT		1,706 (47.88%)		1,653 (48.09%)		1,842 (50.17%)

Table 14: List of the 10 most recurrent lemmatised verbs in the three subcorpora.

Taking a few examples from the verb *do* used as a main verb, a few differences can be identified in its use. Whereas in most sentences national speakers use the verb in concise and emphatic expressions or fixed expressions (e.g. *do your best*), in the other two subcorpora there are several examples in which the verb seems to be used owing to a lack of a more precise alternative or in a moment in which the speaker is stalling (as can be inferred from repetitions). More specifically, in the case of interpreters, in a few examples the verbs *do* are redundant as they do not convey additional meaning (e.g. INT (7), below) and are possibly used to buy time or as an approximation strategy¹⁴ (see Kohn / Kalina 1996).

NAT (4)

I know you're *doing* your best. (N __ S1.1)

NAT (5)

Whatever we *do*, we actually change the system of payment and actually move farmers in another direction. (N __ S7)

NAT (6)

Secondly, talk about the opportunity to *do* things differently. (N __ S9)

EUR (7)

We want to be an independent, self-governing, normal nation. And that is what we have *done* and that is what must happen. And in *doing* so / and in *doing* so, we now offer a beacon of hope. (EUR __ S3)

EUR (8)

So, the question / the question is, what do we *do* next? (EUR __ S3)

¹⁴ Approximating consists in paraphrasing or using an approximate translation when the interpreter cannot access the ideal translation due to time pressure.

EUR (9)

First of all, we should continue to *do* while we currently *do* well. (EUR __ S1)

INT (4)

That's what we've *done* for the citizens because what we say we *do*. (INT __ S5)

INT (5)

Everything's been *done* very quickly. (INT __ S14)

INT (6)

We will *do* this with the support and consensus of the people of Italy.

(INT __ S6)

INT (7)

And that's a personal view that I have I think what we have to try and avoid *doing* in Europe is to follow the Trump model. (INT __ S8)

INT (8)

It's a negative idea. Why? Because what it *does* is it breaks the historical contract that we have. (INT __ S8)

The same tendency registered for verbs can be observed for adverbs as fewer items (194) are used in comparison to the other two subcorpora (201 for the national and 205 for the European), even though adverbs account for a larger share of lexical words in the EN interpreted subcorpus (11.44% as opposed to 10.75% for the national and 10.16% for the European). Table 15 shows the adverbs' high frequency list – that is the first ten most recurrent lemmatised adverbs – in the three subcorpora. In this case too, even though there is a major overlap as to the most recurrent adverbs (as is the case for verbs), the first ten adverbs account for 50.19% of all the adverbs used by interpreters (against 44.39% in the case of the EN national and 46.21% in the case of the EN European).

	EN NAT		EN EUR		EN INT	
1	not	164	not	121	not	180
2	so	85	so	110	so	101
3	also	67	very	55	now	67
4	very	51	also	51	also	67
5	as	40	now	45	very	60
6	actually	27	well	36	as	41
7	just	27	just	31	well	39
8	now	25	as	29	just	36
9	much	25	more	28	here	33
10	then	24	here	25	then	24
TOT		535 (44.39%)		531 (46.21%)		648 (50.19%)

Table 15: List of the 10 most recurrent lemmatised adverbs in the three subcorpora.

For the category of adjectives (see Table 11), there is a less evident tendency, and the outlier is rather the EN European subcorpus with both a higher number of single items (477 as opposed to 448 for the EN national and 447 for the EN interpreted) and a higher total frequency (14.18% of the total lexical words as opposed to 13.88% for the EN national and 14.02% for the EN interpreted). The data on the ten most recurrent lemmatised adjectives (see Table 16) confirm that the EN European subcorpus is the least repetitive as the adjectives' list head is responsible for only 18.76% of adjectives (against the 22.62% in the EN national and 21.82% in the EN interpreted). Even though the lists are very similar in terms of which the most recurrent adjectives are, the European dimension is considerably more prominent in the two subcorpora addressed to an international audience (32 occurrences in the EN national, 63 in the EN European and 78 in the EN interpreted), thus confirming that, irrespective of the topics addressed, the audience plays a pivotal role in the framing of the themes and the relevance attributed to underlying recurrent topics, as is the case of the European Union and the European dimension.

	EN NAT		EN EUR		EN INT	
1	new	49	european	63	european	78
2	other	48	good	35	important	52
3	important	42	new	35	first	36
4	good	37	many	32	italian	31
5	agricultural	36	last	31	other	29
6	more	32	more	29	national	28
7	european	32	other	27	new	23
8	many	27	important	26	economic	23
9	common	26	political	24	unfair	23
10	clear	23	first	21	clear	22
TOT		352 (22.62%)		301 (18.76%)		345 (21.82%)

Table 16: List of the 10 most recurrent lemmatised adjectives in the three subcorpora.

4.4 INTERPRETESE AND NATIVE LANGUAGE USE OF FIGURES OF SPEECH

In this part of the research, an analysis on the frequency of use of three figures of speech was carried out, namely the anaphora, the tricolon and the antithesis. Figures of speech have been chosen as the subject matter of this part of the study because of their great importance and extensive use in political discourse; Eco argues that a figure of speech enables one to say something that may already be known in a new and surprising way, thus capturing the listen-

er's attention (1987: 95). More specifically regarding political discourse, Teun A. van Dijk (1998: 25) argues that

lexical items not only may be selected because of official criteria of decorum, but also because they effectively emphasize or de-emphasize political attitudes and opinions, garner support, manipulate public opinion, manufacture political consent, or legitimate political power. The same may be true for the selection of topics, for the use of rhetoric figures, the pragmatic management of speech acts, interactional self-presentation, and so on.

The following are definitions provided by scholars of the three figures of speech identified in the three subcorpora.

Harris defines the anaphora as “the repetition of the same word or words at the beginning of successive phrases, clauses, or sentences, commonly in conjunction with climax and with parallelism” (1997: 17).

An example of this from the EN national subcorpus is the following:

NAT (7)

More action to bring dwindling fish stocks back to our rivers. *More action* to save other endangered species from pangolins to rhinos. *More action* to develop the technologies which will free us from reliance on harmful chemicals. *More action* to help our bees and pollinators (N __ S3)

The tricolon creates an equivalence in the syntactic structure, words, phrases or sentences of the speaker (Mortara Garavelli 1988: 230). This figure of speech is usually used to create a rhetorical effect, to capture the attention of the listeners and to be easily remembered by them.

An example of this from the EN national subcorpus is the following:

NAT (8)

And the European Council's Guidelines aspire to a *balanced, ambitious, and wide-ranging* deal, with common rules in a number of areas to ensure fair and open competition (N __ S1.1)

Lastly, the antithesis is defined by Harris as a rhetorical figure that “establishes a clear, contrasting relationship between two ideas by joining them together or juxtaposing them, often in parallel structure” (Harris 1997: 16).

An example of antithesis from the EN national subcorpus is the following:

NAT (9)

The government I lead will be driven not by the interests of the privileged few, but by yours (N __ S1.1)

These particular figures of speech were chosen following a corpus-driven approach (see Introduction). By studying the corpus, it quickly became appar-

ent that these were the most common figures of speech and were therefore selected for this part of the analysis. The aim of this analysis was to determine whether there were remarkable differences in the frequency of use of figures of speech between the speakers from the three subcorpora, i.e. EN national, EN European and EN interpreted and, if so, postulate the reasons underlying such differences.

In some cases speakers used the anaphora and the tricolon together, in this case both the anaphora and the tricolon were counted. An example of this overlap from the EN European subcorpus is the following:

EUR (10)

They rejected the multinationals. They rejected the merchant banks. They rejected big politics (EUR_S3)

Tables 17, 18 and 19 indicate the single speakers from each subcorpus, along with the number of instances in which they employed anaphoras, tricolons and antitheses.

EN national	Length in words	Anaphoras	Tricolons	Antitheses	
N_S1.1	6,815	4	9	10	
N_S1.2	864	1	0	0	
N_S2	77	0	0	0	
N_S3	2,272	9	13	1	
N_S4	2,393	1	0	0	
N_S5	743	1	2	0	
N_S6	854	0	1	0	
N_S7	1,532	3	1	1	
N_S8	889	1	1	1	
N_S9	870	1	1	1	
N_S10	1,187	0	0	0	
N_S11	722	0	1	0	
N_S12	981	0	0	0	
		21	29	14	Total 63

Table 17: List of figures of speech used in the EN national subcorpus.

EN European	Length in words	Anaphoras	Tricolons	Antitheses	
EUR__S1	3,234	0	7	0	
EUR__S2	674	1	1	0	
EUR__S3	755	4	4	1	
EUR__S4.1	4,283	2	6	2	
EUR__S5	358	0	0	0	
EUR__S4.2	2,281	0	3	1	
EUR__S6	1,869	0	3	2	
EUR__S4.3	1,776	0	2	0	
EUR__S7	1,742	0	1	0	
EUR__S8	640	0	2	1	
EUR__S9	771	0	0	1	
EUR__S10	1,646	1	3	2	
		8	32	10	Total 50

Table 18: List of figures of speech used in the EN European subcorpus.

EN interpreted	Length in words	Anaphoras	Tricolons	Antitheses	
INT__S1.1	4,419	1	3	1	
INT__S2	424	0	0	0	
INT__S3	610	0	0	1	
INT__S4	880	0	1	0	
INT__S5	676	0	4	1	
INT__S6	458	2	2	0	
INT__S7	223	1	0	0	
INT__S1.2	2,285	2	3	1	
INT__S8	2,197	1	1	2	
INT__S9	139	0	0	0	
INT__S10	1,914	0	3	0	
INT__S11	139	0	0	0	
INT__S12	1,974	0	2	0	
INT__S13	83	0	0	0	
INT__S14	981	2	0	0	
INT__S15	1,169	0	1	0	
INT__S16	1,155	0	3	0	
		9	23	6	Total 38

Table 19: List of figures of speech used in the EN interpreted subcorpus.

The results show an overall higher number of figures of speech used in the EN national subcorpus (63), compared to the EN European subcorpus (50) and to the EN interpreted subcorpus (38). Furthermore, the most often used figure of speech in all three subcorpora is the tricolon. Instead, the least used figure of speech is in two cases the antithesis (EN national and EN interpreted) and in one case the anaphora (EN European). However, the figures of speech of the EN national subcorpus are high largely owing to two speeches in particular, namely N__S1.1, the 2018 Mansion House speech and N__S3, the 2018 speech at the Conservative Party Conference: both are ceremonial speeches where more sophisticated, figurative language solutions may be expected and were indeed identified. Without these speeches, which constitute together 70.76% of all figures of speech in the EN national subcorpus, the number of figures of speech from the EN national subcorpus and the other subcorpora would be more similar. Furthermore, by observing the figures of speech present in EUR__S3 in the EN European subcorpus, one can note the following: although it was only 5:57 minutes long, with a total of 755 words in total, EUR__S3 includes 4 anaphoras, 4 tricolons and 1 antithesis. In EUR__S3 the speaker therefore employed more figures of speech than most other speakers in the EN European subcorpus despite speaking for less time in comparison with many other speakers. This may be an indication that the presence of figures of speech or lack thereof in a given speech may depend more on the type of event taking place (see N__S1.1, the Mansion House speech, and N__S3 delivered at the Conservative Party Conference) and on the personal style of the speaker delivering the speech (see EUR__S3 at the European Parliament), rather than on whether the speech is addressed to a national or international audience.

5. CONCLUSIONS

The results obtained from this study are not fully in line with translation universals. In terms of generalisation, the EN interpreted subcorpus is the most lexically dense (highest proportion of lexical words to function words), thus making it the most informative of the three. A possible explanation could be that interpreters are taught not to be redundant, to identify repetitions in the source text – which are not used as rhetorical devices – and leave them out of their target text. However, the EN interpreted subcorpus presents a higher degree of repetitiveness precisely as concerns lexical words. As shown in the case of the verb to *do*, a possible explanation could be that interpreters do not have time to look for synonyms or reformulations and once they have found a suitable equivalent for a concept expressed in a speech they tend to stick to it, which would make their speeches highly informative but less varied in terms of style and vocabulary. Indeed, in all lexical groups (with the exception of verbs) they presented the lowest number of items per category.

The analysis also showed that the interpreters resorted to the optional *that*

less than the speakers from the other subcorpora, which is not in line with the interpreters' explicitness witnessed in previous studies. This result may be a consequence of the interpreters applying specific interpreting strategies, for example in this case avoiding the use of non-essential words.

Furthermore, some differences were also observed in the use of figures of speech, which may be due to the type of event taking place and to the speaker's personal style, rather than to the speech being addressed to a national or international audience.

Although the analysis is carried out on a relatively small corpus and is limited in its scope since it focuses on a selection of traits, the differing results from the three subcorpora can nevertheless be a positive incentive to carry out similar research on larger corpora of this type in the future.

REFERENCES

- Baker M. (1996) "Corpus-based Translation Studies: The challenges that lie ahead", in H. Somers (ed.) *Terminology, LSP and Translation: Studies in Language Engineering in Honour of Juan C. Sager*, Amsterdam/Philadelphia, John Benjamins, 175-178.
- Baker M. (1993) "Corpus linguistics and translation studies – Implications and applications", in M. Baker / G. Francis / E. Tognini-Bonelli (eds) *Text and Technology*, Amsterdam/Philadelphia, John Benjamins, 233-252.
- Chesterman A. (2004) "Beyond the particular", in A. Chesterman / P. Kujamäki (eds) *Translation Universals. Do they exist?*, Amsterdam/Philadelphia, John Benjamins, 33-49.
- Eco, Umberto (1987) "Il linguaggio politico", in G. Beccaria (ed.) *I linguaggi settoriali in Italia*. Milano, Bompiani. 91-105.
- Ferraresi A. / Milicevic M. (2017) "Phraseological patterns in interpreting and translation. Similar or different?", in G. De Sutter / M.A. Lefer / I. Delaere (eds), *Empirical Translation Studies: New Methodological and Theoretical Traditions*, Berlin/Boston, De Gruyter Mouton, 1-30.
- Harris R. (1997) *A Handbook of Rhetorical Device*, <<https://hellesdon.org/documents/Advanced%20Rhetoric.pdf>>.
- Jurafsky D. / Martin J. H. (2004) "Word classes and part-of-speech tagging", revised 2004 version, original chapter in *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, Upper Saddle River, Prentice Hall, 2000, 1-54.
- Kajzer-Wietrzny M. (2012) *Interpreting universals and interpreting style*, PhD Thesis, Poznań, AMU, Adam Mickiewicz University.
- Kohn K. / Kalina S. (1996) "The strategic dimension of interpreting", *Meta* 41/1, 118-138.
- Laviosa S. (1998) "Core patterns of lexical use in a comparable corpus of English narrative prose", *Meta* 43/4, 557-570.
- McEnery T. / Hardie A. (2011) *Corpus Linguistics: Method, Theory and Practice*, Cambridge, Cambridge University Press.
- McEnery T. / Xiao R. / Tono Y. (eds) (2005) *Corpus-Based Language Studies. An Advanced Resource Book*. London/New York, Routledge.
- Mortara Garavelli B. (1988) *Manuale di retorica*, Milano, Bompiani.
- Olohan M. (2002) "Corpus Linguistics and Translation Studies: Interaction and reaction", *Linguistica Antverpiensia* 1, 419-429.
- Olohan M. / Baker M. (2000) "Reporting that in translated English. Evidence for subconscious processes of explicitation?", *Across Languages and Cultures* 1/2, 141-158.
- Procter P. (1995) *Cambridge International Dictionary of English*, Cambridge, Cambridge University Press.
- Sandrelli A. / Bendazzoli C. (2005) "Lexical patterns in simultaneous interpreting: a preliminary investigation of EPIC (European Parliament Interpreting Corpus)", *Proceedings from the Corpus Linguistics Conference Series*, 1/1, <<https://www.birmingham.ac.uk/research/activity/corpus/publications/conference-archives/2005-conf-e-journal.aspx>>.
- Shlesinger M. (2009) "Towards a definition of interpretese: an intermodal, corpus-based study", in G. Hansen / A. Chesterman / H. Gerzymisch-Arbogast (eds) *Efforts and Models in Interpreting and Translation Research: A Tribute to Daniel Gile*, Amsterdam/

Philadelphia, John Benjamins, 237-253.

Shlesinger M. (1998) "Corpus-based Interpreting Studies as an offshoot of Corpus-based Translation Studies", *Meta* 43/4, 486-493.

Shlesinger M. / Ordan N. (2012) "More spoken or more translated? Exploring a known unknown of simultaneous interpreting", *Target* 24/1, 43-60.

Stubbs M. (1996) "Text and Corpus Analysis: Computer-assisted studies of language and culture", *Functions of Language* 3/2, 269 - 272.

Tognini-Bonelli E. (2001) *Corpus Linguistics at Work*, Amsterdam/Philadelphia, John Benjamins.

Toury G. (2004) "Probabilistic explanations in translation studies: Welcome as they are, would they qualify as universals?" in A. Chesterman / P. Kujamäki (eds) *Translation Universals. Do they exist?*, Amsterdam/Philadelphia, John Benjamins, 15-32.

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

van Dijk T.A. (1998) "What is Political Discourse Analysis?", *Belgian Journal of Linguistics* 11/1, 11-52.

Zanettin F. (2012) *Translation-driven Corpora. Corpus Resources for Descriptive and Applied Translation Studies*, Manchester, St. Jerome.