

Marta Kajzer-Wietrzny

Interpreting universals and interpreting style

Rozprawa doktorska napisana
na Wydziale Anglistyki
Uniwersytetu im. Adama Mickiewicza
pod kierunkiem
Prof. UAM dr hab. Małgorzaty Fabiszak

Poznań, 2012

OŚWIADCZENIE

Ja, niżej podpisana

Marta Kajzer-Wietrzny

studentka Studium Doktoranckiego na Wydziale Anglistyki

Uniwersytetu im. Adama Mickiewicza w Poznaniu

oświadczam,

że przedkładaną rozprawę doktorską

pt. *Interpreting universals and interpreting style*

napisałam samodzielnie.

Oznacza to, że przy pisaniu pracy, poza niezbędnymi konsultacjami, nie korzystałam z pomocy innych osób, a w szczególności nie zlecałam opracowania rozprawy lub jej istotnych części innym osobom, ani nie odpisywałam tej rozprawy lub jej istotnych części od innych osób.

Jednocześnie przyjmuję do wiadomości, że gdyby powyższe oświadczenie okazało się nieprawdziwe, decyzja o wydaniu mi dyplomu zostanie cofnięta.

Poznań, 13.04.2012r.

(miejsowość, data)

(czytelny podpis)

Acknowledgements

I would like to express sincere gratitude to my supervisor, Professor Małgorzata Fabiszak, for invaluable constructive criticism, stimulating discussions, great attention to detail and immense support extending far beyond her formal responsibilities.

I wish to extend my thanks to my former supervisor, Professor Alicja Pisarska, who provided careful guidance at the initial stages of my academic endeavours as a PhD student.

I am very grateful to Professor Miriam Shlesinger for very a fruitful discussion on my project and the patience that she exercised when listening to my ideas.

I am also greatly indebted to Agnieszka Chmiel and Przemysław Kaszubski for assisting me with their expertise on countless occasions.

Special thanks are due to Professor Homayoon Beigi and Recognition Technologies, Inc. for their support given in the speaker recognition process.

I am grateful to the Polish Ministry of Education for financing this project (grant No. N N104 053239).

Finally, I would like to thank my husband and my parents for their encouragement and support.

Table of contents

TABLE OF CONTENTS	4
LIST OF TABLES	7
LIST OF FIGURES	12
INTRODUCTION	13
CHAPTER 1: UNIVERSALS AND STYLE IN CORPUS-BASED TRANSLATION AND INTERPRETING STUDIES	17
1.1. UNIVERSAL AND INDIVIDUAL TENDENCIES.....	17
1.2. FROM DESCRIPTIVE TRANSLATION STUDIES TO CORPUS-BASED TRANSLATION STUDIES	18
1.3. CORPUS TYPOLOGY AND CORPUS PROCESSING METHODS	20
1.4. UNIVERSALS AS A MAJOR RESEARCH AREA IN CORPUS-BASED TRANSLATION STUDIES	23
1.4.1. <i>Simplification</i>	26
1.4.2. <i>Explicitation and explicitness</i>	27
1.4.3. <i>Normalisation</i>	34
1.4.4. <i>Levelling out</i>	36
1.4.5. <i>Interference</i>	36
1.5. TRANSLATION UNIVERSALS REFLECTED IN SELECTED INTERPRETING RESEARCH ..	38
1.5.1. <i>Simplification</i>	38
1.5.2. <i>Explicitation</i>	40

1.5.3. <i>Normalisation</i>	42
1.5.4. <i>Levelling-out or convergence</i>	42
1.6. CORPUS-BASED INVESTIGATIONS OF TRANSLATOR'S STYLE.....	43
1.7. CONCLUSIONS	47
CHAPTER 2: OBJECTIVES AND METHODS	49
2.1. OBJECTIVES	49
2.1.1. <i>Operationalisation of research questions</i>	51
2.2. DATA	57
2.2.1. <i>Translation and Interpreting Corpus (TIC)</i>	57
2.2.2. <i>Individual interpreting corpora (parts of TIC)</i>	71
2.2.3. <i>Reference corpora of interpreters' spoken discourse</i>	72
2.3. METHOD FOR MACRO ANALYSIS: INTERPRETING UNIVERSALS	73
2.3.1. <i>Simplification</i>	74
2.3.2. <i>Explicitness</i>	75
2.3.3. <i>Normalisation</i>	77
2.4. METHOD FOR MICRO ANALYSIS: INTERPRETING STYLE	81
2.4.1. <i>Simplification</i>	82
2.4.2. <i>Explicitness</i>	83
2.4.3. <i>Normalisation</i>	84
2.5. MEASURE OF STATISTICAL SIGNIFICANCE	84
2.6. LIMITATIONS OF THE APPLIED METHODOLOGY	85
CHAPTER 3: MACRO ANALYSIS OF RECURRENT PATTERNS IN INTERPRETING.....	88
3.1. INTRODUCTION	88
3.1.1. <i>Simplification</i>	89
3.1.2. <i>Explicitness</i>	98
3.1.3. <i>Normalisation</i>	108
3.2. OBSERVED PATTERNS, THEIR UNIVERSALITY AND IDIOSYNCRATIC STYLISTIC VARIATION.....	114
CHAPTER 4: MICRO ANALYSIS OF IDIOSYNCRATIC VARIATION: INTERPRETING STYLE	117

4.1. INTRODUCTION	117
4.2. NON-LINGUISTIC DATA FROM INTROSPECTION	118
4.2.1. <i>Simplification</i>	119
4.3. EXPLICITNESS	125
4.3.1. <i>Optional that</i>	126
4.4. NORMALISATION.....	129
4.4.1. <i>Fixed expressions</i>	129
4.5. SIMPLIFICATION, EXPLICITNESS AND NORMALISATION AS IDIOSYNCRATIC STYLISTIC FEATURES	131
4.6. STYLE VS. UNIVERSAL TENDENCIES	132
4.7. LIMITATIONS OF THE MICRO STUDY OF IDIOSYNCRATIC VARIATION: INTERPRETING STYLE.....	134
CONCLUSIONS	136
REFERENCES.....	151
APPENDICES	160
ABSTRACT IN POLISH	165

List of tables

Table 1. Transcription conventions in TIC adapted from EPIC.....	70
Table 2. Trigrams selected for investigation.	78
Table 3. The number of tokens in the list head vs. the number of remaining tokens in ORG_SP_EN and SI_FR_EN.....	90
Table 4. Proportion of the corpus covered by list head in TIC oral subcorpora (the higher the values, the more repetitive i.e. simplified the text).	90
Table 5. Differences between oral corpora of TIC with respect to list head proportion: results of Chi-square tests calculated on raw counts in 2x2 contingency tables.....	90
Table 6. Proportion of the corpus covered by list head in the subcorpora of native English speeches, interpretations of Spanish speeches that were read out and interpretations of Spanish speeches delivered orally (the higher the values, the more repetitive i.e. simplified the text).	91
Table 7. Statistical significance of differences in the proportion of the corpus covered by list head in the subcorpora of native English speeches, interpretations of Spanish speeches that were read out and interpretations of Spanish speeches delivered orally.	91
Table 8. Lexical density in TIC oral subcorpora (the lower the values, the less lexically dense i.e. simplified the text).	93
Table 9. Differences between oral corpora of TIC with respect to lexical density: results of Chi-square tests calculated on raw counts in 2x2 contingency tables.	93
Table 10. Lexical density in the subcorpora of native English speeches, interpretations of Spanish speeches that were read out and interpretations of Spanish speeches delivered orally (the lower the values, the more simplified the text).	94

Table 11. Statistical significance of differences in lexical density in the subcorpora of native English speeches, interpretations of Spanish speeches that were read out and interpretations of Spanish speeches delivered orally.	94
Table 12. The proportion of high frequency words in TIC oral subcorpora (the higher the values, the narrower the range of vocabulary i.e. simplified the text).	96
Table 13. Differences between oral corpora of TIC with respect to the proportion of high frequency words: results of Chi-square tests calculated on raw counts in 2x2 contingency tables.	96
Table 14. High frequency words in the subcorpora of native English speeches, interpretations of Spanish speeches that were read out and interpretations of Spanish speeches delivered orally (the higher the values, the more simplified the text).	97
Table 15. Statistical significance of differences in the proportion of high frequency words in the subcorpora of native English speeches, interpretations of Spanish speeches that were read out and interpretations of Spanish speeches delivered orally.	97
Table 16. Occurrences of optional <i>that</i> and zero-connectives after reporting verbs in TIC spoken subcorpora (z.c. stands for zero connective).	99
Table 17. Differences between oral corpora of TIC with respect to occurrences of optional <i>that</i> : results of Chi-square tests calculated on raw counts.	99
Table 18. Occurrences of optional <i>that</i> and zero-connectives in TIC written subcorpora, values normalised per 50000 and rounded (z.c. stands for zero connective).	100
Table 19. Differences between oral corpora of TIC with respect to occurrences of optional <i>that</i> : results of Chi-square tests calculated on raw counts.	100
Table 20. Occurrences of linking adverbials in TIC spoken subcorpora.	104
Table 21. Differences between oral corpora of TIC with respect to the occurrences of linking adverbials: results of Chi-square tests calculated on raw counts in 2x2 contingency tables.	104
Table 22. Occurrences of linking adverbials in TIC written subcorpora normalised per 50000 words.	104
Table 23. Differences between oral corpora of TIC with respect to the occurrences of linking adverbials: results of Chi-square tests calculated on raw counts in 2x2 contingency tables.	104
Table 24. Occurrences of selected apposition markers in TIC spoken subcorpora.	106

Table 25. Differences between oral corpora of TIC with respect to the occurrences of apposition markers: results of Chi-square tests calculated on raw counts in 2x2 contingency tables.	106
Table 26. Occurrences of <i>in other words</i> in Dutch interpreting corpus broken down by interpreters.....	107
Table 27. Occurrences of selected apposition markers in TIC written subcorpora. Values normalised per 50000 words and rounded to the full digit.	107
Table 28. Differences between written corpora of TIC with respect to the occurrences of apposition markers: results of Chi-square tests calculated on raw counts in 2x2 contingency tables.	107
Table 29. Lexical bundles in TIC spoken subcorpora.	110
Table 30. Differences between oral corpora of TIC with respect to the occurrences of lexical bundles: results of Chi-square tests calculated on raw counts in 2x2 contingency tables.	110
Table 31. Lexical bundles in TIC written subcorpora. Frequencies are normalised per 50000 words and rounded to the full digits.....	112
Table 32. Differences between written corpora of TIC with respect to the occurrences of lexical bundles: results of Chi-square tests calculated on raw counts in 2x2 contingency tables.	112
Table 33. Occurrences of the most frequent fixed expressions in TIC spoken subcorpora.	113
Table 34. Differences between oral corpora of TIC with respect to the occurrences of most frequent fixed expressions: results of Chi-square tests calculated on raw counts in 2x2 contingency tables.	113
Table 35. Occurrences of the most frequent fixed expressions in TIC written subcorpora Frequencies are normalised per 50000 words and rounded to the full digits.....	113
Table 36 Differences between written corpora of TIC with respect to the occurrences of most frequent fixed expressions: results of Chi-square tests calculated on raw counts in 2x2 contingency tables.	113
Table 37. Juxtaposition of the outcomes of macro analysis.	116
Table 38. Summary of the answers given by interpreters in the warm-up session.....	118

Table 39. Proportion of the corpus covered by list head analysed on 720 word corpora of interpreter I1's non-interpreted and interpreted discourse consisting of text samples.	120
Table 40. Repetitiveness of I1's interpreted and non-interpreted discourse: Results of Chi-square tests calculated on raw counts in 2x2 contingency tables.....	120
Table 41. Proportion of the corpus covered by list head analysed on 720 word corpora of interpreter I2's non-interpreted and interpreted discourse consisting of text samples.	121
Table 42. Repetitiveness of I2's interpreted and non-interpreted discourse: Results of Chi-square tests calculated on raw counts in 2x2 contingency tables.....	121
Table 43. Repetitiveness of I1 vs. I2: Results of Chi-square tests calculated on raw counts in 2x2 contingency tables.....	122
Table 44. Lexical density analysed on 720 word corpora of interpreter I1's non-interpreted and interpreted discourse consisting of text samples.	123
Table 45. Lexical density of I1: Results of Chi-square tests calculated on raw counts in 2x2 contingency tables.	123
Table 46. Lexical density analysed on 720 word corpora of interpreter I2's non-interpreted and interpreted discourse consisting of text samples.	123
Table 47. Lexical density of I2: Results of Chi-square tests calculated on raw counts in 2x2 contingency tables.	123
Table 48. The proportion of high frequency words analysed on 720 word corpora of interpreter I1's non-interpreted and interpreted discourse consisting of samples..	124
Table 49. Lexical sophistication of I1's interpreted and non-interpreted discourse: Results of Chi-square tests calculated on raw counts in 2x2 contingency tables.	124
Table 50. The proportion of high frequency words analysed on 720 word corpora of interpreter I2's non-interpreted and interpreted discourse consisting of text samples.	125
Table 51. Lexical sophistication of I1's interpreted and non-interpreted discourse: Results of Chi-square tests calculated on raw counts in 2x2 contingency tables.	125
Table 52. The proportion of optional <i>that</i> after reporting verbs analysed on the full collection of interpreter I1's interpretations and the full transcript of non-interpreted discourse normalised per 1000 words. Raw frequencies in brackets.....	126

Table 53. The proportion of optional <i>that</i> after reporting verbs analysed on the full collection of interpreter I2's interpretations and the full transcript of non-interpreted discourse normalised per 1000 words. Raw frequencies in brackets.....	128
Table 54. The frequency of fixed expressions in I1's interpreted and non-interpreted discourse normalised per 1000 words.....	130
Table 55. The frequency of fixed expressions in I2's interpreted and non-interpreted discourse normalised per 1000 words.....	130
Table 56. Juxtaposition of the outcomes of micro analysis.....	135
Table 57. Juxtaposition of the outcomes of macro and micro analysis.....	137

List of figures

Figure 1. The structure of macro analysis of interpreting universals depicting the investigated features, parameters and data sets.	54
Figure 2. The structure of micro analysis of interpreting style depicting the investigated features, parameters and data sets.	56
Figure 3. Translation and Interpreting Corpus.	58
Figure 4. Composition of oral subcorpora of TIC.	59
Figure 5. Composition of written subcorpora of TIC.	60
Figure 6. The distribution of interpretations carried out by individual interpreters across respective subcorpora.	63
Figure 7. The total length of texts interpreted by individual interpreters from different languages expressed in number of words.	64
Figure 8. Topics in oral subcorpora of TIC.	65
Figure 9. Mode of delivery of original speeches across respective interpreting subcorpora (expressed in number of tokens).	66
Figure 10. Topics in written subcorpora of TIC.	67
Figure 11. Location of TIC subcorpora on the literacy – orality axis.	68
Figure 12. Mode of delivery of original speeches across respective corpora of Interpreter 1’s interpretation (expressed in number of tokens).	71
Figure 13. Mode of delivery of original speeches across respective corpora of Interpreter 2’s interpretation (expressed in number of tokens).	72

Introduction

The unique nature of translated texts has for decades now been the focal point of Descriptive Translation Studies. The phenomenon is so evident that it has received labels, such as *translationese* (Toury 1980, Gellerstam 1986) or *the third code* (Frawley 1984), and yet so elusive that researchers are still struggling to unveil all its characteristic traits.

The search for features that typify translations was at the beginning based on micro analyses of different linguistic and paralinguistic strata of single texts and their translations. Most conspicuous ones such as the tendency to use more simplified, explicit and conventional language were early revealed in a number of independent studies. Once the preliminary hypotheses were formulated, time came for a great methodological advancement – corpus linguistics that allowed to automatically analyse large quantities of texts available in machine readable form. In the years to follow, researchers went to great lengths to devise appropriate methods which would help them test old hypotheses with the new tool. Having analysed a multitude of translations in different language pairs scholars hailed a set of the recurrent features translation universals.

While hypotheses regarding universals are still being refined, a new trend has emerged focusing on the identification of idiosyncratic stylistic preferences of individual translators, in other words, translator style, also with the use of corpus linguistics methods. Although in its early days the investigation of translator style focused mostly on discerning features characteristic of individual translators that make them distinct from others, the methodology may also be used to corroborate or refute translation universals and verify if they are more prominent than one's idiosyncratic stylistic preferences.

Although still very unstable in its foundations and lacking conceptual clarity, the notion of translation universals is a recognised and actively developing branch of

translation studies relying now very heavily on corpus linguistics methods. The same cannot be said, however, about Interpreting Studies, where corpus based research is still in its infancy. Most reports on interpreting universals are based on single cases or small sized corpora. Focusing on the micro analysis researchers rarely reach for methods that have been developed to research translation universals, which makes the findings of both sibling disciplines difficult to compare. There are notable exceptions to this rule, like the investigation of simplification on the European Parliament Interpreting Corpus (Sandrelli – Bendazolli 2005), but they are few and far between. The phenomenon of interpreting style i.e. interpreter's idiosyncratic stylistic preferences has not been explored with corpus linguistics methods at all. This gap can most probably be explained by the fact that such studies are difficult to carry out on a large scale. Authentic interpreting data is not easily accessible and compilation of a corpus requires meticulous and time consuming transcription work.

The overriding aim of the present dissertation is to, at least in part, fill this lacuna by an attempt to verify if three features hailed to be translation universals: simplification, explicitness and normalisation typify also interpreted texts and to what extent their prominence is dependent on interpreters' idiosyncratic stylistic preferences i.e. interpreting style.

The hypothesised universality of the three investigated features will be tested with a set of parameters developed in translation studies research on Translation and Interpreting Corpus (TIC) comprising subcorpora of interpretations and translations into English from four different languages: French, Spanish, German and Dutch as well as speeches produced originally in English (both oral and written) used as reference corpora. This set of procedures will allow to verify on a macro scale whether examined features are plausible interpreting universals i.e. if they are consistently more prominent in all subcorpora comprising interpretations from different languages into English than in the subcorpus of speeches originally produced in English. Once the global recurring patterns are established, micro analysis will follow aimed at investigating stylistic preferences of two interpreters. Micro analysis will be conducted on small corpora of texts interpreted into English by two interpreters from different source languages and compared to small corpora of their own non-interpreted discourse. This part of the study will allow to determine, if global patterns are equally reflected in the interpreting output of individual interpreters, or whether their

stylistic preferences turn out to be more prominent than global tendencies and consistent enough to point to characteristic features of their distinctive interpreting style.

Chapter One of the present dissertation reviews the development of the quest for translation universals focusing mostly on the achievements of corpus based studies in this area. It analyzes briefly major corpus linguistics methods and tools most frequently applied in translation studies research, reports on seminal studies in corpus based analyses devoted to translation universals and relevant investigations that have been conducted in the field of interpreting. The presentation of key research in the area ends with a report on most important corpus based analyses of translator style.

Chapter Two presents in greater detail the objectives of this study briefly indicated above and discusses the exact procedures of macro analysis focused on interpreting universals and micro analysis examining the intricacies of interpreting style. It also provides a comprehensive description of data used in both analyses, discusses the compilation of parameters applied to the investigation of simplification, explicitness and normalisation and statistical tests utilised in this dissertation.

Chapter Three reports the outcome of the macro analysis of interpreting universals carried out on the Translation and Interpreting Corpus comprising interpretations and translations from four different languages into English. The focal point of the macro analysis is to determine if the examined features are evenly distributed across all interpreting corpora and to what extent the translation universals are reflected in the language of interpretation.

Chapter Four discusses the findings of micro analysis of interpreting style, in which the interpreted and non-interpreted discourse of two individual interpreters is compared. The key areas of investigation include the differences between the language used by both interpreters and the general patterns observed in TIC, differences in their interpretations from Romance and Germanic languages and the language they use while speaking.

The outcomes of macro and micro analyses are juxtaposed and discussed in a summarised way in Conclusions, where areas that need further investigation are also indicated.

It is hoped that thanks to the innovative, broad perspective that joins the investigations of universals with the analysis of style, the study described in this thesis will help expand the current knowledge on the nature of interpreted language, its recurring

features as well as several factors that may condition their prominence such as the source language, mode of delivery of the original speech and the interpreters' stylistic preferences.

Chapter 1: Universals and style in Corpus-based Translation and Interpreting Studies

1.1. Universal and individual tendencies

The quest for universals in corpus-based translation research has rarely been combined with the analysis of individual preferences. The two lines of research have so far been pursued parallelly, even though they mutually complement each other. There is a clear benefit in investigating universal patterns in the work of individuals, as the universality of the analysed feature is questionable, if idiosyncratic features of an individual overshadow the hypothetically universal ones to the point of eliminating them entirely. On the other hand, the analysis of style may also gain a new perspective, if one's idiosyncratic traits could be juxtaposed with general trends. This, however, is rarely done.

While the investigations of translation universals and translator's style follow parallel tracks, they are both gaining a steady ground in the field of corpus-based research. Compared to what has already been achieved in research on translation, corpus-based research devoted to interpreting is still in the early phase of development: large-scale studies examining recurrent patterns in interpreting are scarce and the phenomenon of interpreting style has not yet been explored at all.

This Chapter reviews what has so far been achieved in Corpus-based Translation Studies focused on universals and style and to what extent these concepts have been investigated in interpreting research. It sets off with placing Corpus-based Translation Studies within the tradition of Descriptive Translation Studies, presents briefly corpus

typology and corpus methods most frequently used in the investigation of translation universals and style, reviews the key studies on specific translation universals and studies in interpreting research investigating analogical tendencies in this mode. Finally, the last sections in this Chapter summarise corpus-based studies devoted to translator style.

1.2. From Descriptive Translation Studies to Corpus-based Translation Studies

Corpus-based Translation Studies has developed as a product-oriented branch of Descriptive Translation Studies offering a major advancement in methodology allowing to search for recurring patterns and regularities characteristic of translations. Patterns previously observed in individual translations could be tested in an automated manner on a large number of texts tempting researchers to make generalisations.

Descriptive Translation Studies (DTS) can be traced back to the 1970s, when Holmes (1972, 1973, 1988) proposed that Translation Studies should perceive translation in terms of three distinct phenomena: the product, the process and the function. As opposed to the earlier trends in translation research of speculative and prescriptive character, the aim of DTS was to engage with real translation phenomena, to describe and analyse features of translated texts against the contexts in which they were created. Contrary to earlier paradigms focusing on the source text and ideal translation, DTS shifted the focus of research to the target text, which was especially visible in Toury's (1980: 82) claim that translations are facts of target textual tradition only. Focusing on the target text, Toury was inspired by Even-Zohar (1978), who laid foundations for the polysystem theory. The key claim following from his research is an assumption that translated literature constitutes a distinctive entity in the target culture governed by its own system, which dynamically interacts with other systems in the polysystem. As a consequence, translated literature was recognised as an object characterised by specific features worth studying. Following this path, Toury concentrated on the norms i.e. regularities of translational behaviour, which could only be discovered through generalisations made on the basis of a multitude of translations. It soon became clear, however, that Translation Studies lack "strict statistical methods for dealing with translational norms, or even to supply sampling rules for actual research" (Toury 1980: 61). Many problems and doubts of similar nature could be at least partially solved by the application of corpus linguistics methodology in Translation Studies.

This was first suggested by Baker (1993: 243), who in corpus linguistics saw a chance for scholars to uncover the nature of translated text. The new methodology fitted well into the goals of Descriptive Translation Studies allowing researchers to trace regularities and patterns and form generalisations and offering the additional advantage of relatively easy replication. As observed by Laviosa (2008: 122) corpus linguistics makes it possible for DTS to pursue the same comparative research model but expands the scope of examination by increasing the number of texts that could be analysed at once, enabling a unified, consistent and coherent comparison between different text types and even languages and making the generalisations more valid:

Corpus linguistics and Descriptive Translation Studies adapt the same comparative research model in which descriptive hypotheses that make claims about the probabilistic generality of a given phenomenon are put forward and texts are examined across corpora representing different varieties or modalities of the same language (written, spoken and mixed general and specialised monolingual corpora), different languages (bi- and multilingual comparable corpora), translated and non-translated varieties of the same language (monolingual comparable corpora), as well as original texts and their translations into one or more languages (bi- and multilingual parallel corpora) (Laviosa 2008: 122).

Considering all that corpus linguistics methodology has to offer, it is not surprising that soon after Baker's idea was published, researchers set off to test on a larger scale hypotheses and generalisations previously put forward based on analyses of single translations. Such were the cases of Laviosa (1998) investigating the simplification hypothesis initially proposed by Blum-Kulka and Levenston (1983), Toury (1995) or Vanderauwera (1985), or Olohan and Baker (2000) or Mutesayire (2004) examining explicitation first mentioned by Blum-Kulka (1986), to name but a few.

At the turn of the millennium corpus methodology started to be applied also in Interpreting Studies. In her article Shlesinger (1998) analyzes problems and benefits of an application of corpus linguistics methodology and tools in the study of interpreting, which although frequently subsumed under translation, had not until that point been subject to corpus analysis. Considering the fact that interpreting was always studied on the basis of sparse, often anecdotal data Shlesinger (1998: 2) sees a great potential in the "use of large, machine-readable corpora to arrive at global inferences about interpreted text (1) in relation to other forms of oral discourse; and (2) in relation to other forms of translation". She sees, however, two major challenges that researchers have to face: transcription and paralinguistic dimensions. Interpreted text is rarely available in machine-readable form,

therefore compilation of an interpreting corpus is bound to be labour intensive. Following Johansson (1992) she suggests that it might be better to resort to smaller, but carefully constructed sample corpora. The other obstacle is that transcription provides access only to interpreters' linguistic output not reflecting paralinguistic dimensions. Only the advent of new technology may solve this problem. In the meantime researchers try to mitigate concomitant consequences by using a shared system of conventions for encoding auditory data and limit themselves to those features which may be transcribed.

The key area of CTS (Corpus-based Translation Studies) and CIS (Corpus-based Interpreting Studies) research is related to universal features of translation and interpreting, which shall be discussed in greater details in Section 1.4. Corpus-based methodology is, however, successfully applied also in studies of: biblical translation (Masubelele 2003), literary translation (Kruger 2003, Bosseaux 2003), coherence in translation (Behrens – Fabricius-Hansen 2003), translator's style (Baker 2000, Olohan 2004), ideology in translation and interpreting (Kemppanen 2003, Beaton 2007), translation into non-mother tongue (Pokorn 2005), interpreting norms (Niska 2003), legal interpreting (Wallmach 2000, 2003).

1.3. Corpus typology and corpus processing methods

Two major types of corpora are most commonly used in translation research: comparable monolingual corpora and parallel bi- or multilingual corpora. The choice of a corpus depends on the goal of the analysis. Parallel corpora are usually chosen for contrastive analysis and allow to compare the source text and the target text, while comparable corpora are most frequently used in the investigation of recurrent patterns in translated and non-translated language.

Single monolingual corpora contain texts in one language selected according to unified criteria and are usually compared to other monolingual corpora compiled from texts of slightly different characteristics. Corpora might be either translational or non-translational. The latter usually consist of texts originally produced in the target language and serve as reference corpora (to cross-check results of analyses of translational corpora). Translational corpora are examined by scholars focusing on the unique nature of translated text and their characteristic linguistic features e.g. universals or style of individual

translators. Probably the biggest translational corpus is the Translational English Corpus (TEC) at the University of Manchester. It consists of English translations of fiction, biography, news and in-flight magazines from different languages and in 2003 had already reached 10 million words (<http://www.monabaker.com/tsresources/TranslationalEnglishCorpus.htm>) (date of access: 15 November 2011).

If a corpus consists of a set of translated texts and a control set of non-translated texts produced in similar circumstances and serving a similar purpose it is a comparable corpus. Comparable corpora of translational and non-translational language allow for a cross-examination of features of translated and non-translated language usually across a specific text type or genre. Kenny (2001a: 59) provides three examples of such corpora: English Comparable Corpus (ECC) – a sister project of TEC consisting of newspaper articles originally written in English and newspaper articles translated into English; a corpus at the University of Gothenburg comprising 75 novels published in Swedish, approximately half of which are translations into Swedish; a corpus of translational and non-translational Finnish at Savonlinna School of Translation Studies in Joensuu, Finland.

A parallel corpus is usually bilingual, but can also be multilingual. It consists of texts written originally in language A and their translations into other languages, aligned in segments. Alignment may be provided at sentence level, but segments may also be larger and consist of e.g. a whole paragraph. Translation researchers use parallel corpora to investigate lexical correspondence between source and target languages, translation strategies and translation behaviour. Parallel corpora are also used in translation pedagogy or for the development of machine translation. According to Kenny (2001a: 51), parallel corpora have been compiled for several language pairs: English-French, English-Italian, English-Norwegian or English-German. Another important multilingual parallel corpus has been compiled in Italy at University of Bologna, Forlì. European Parliament Interpreting Corpus (EPIC) is a trilingual corpus of Spanish, Italian and English comprising original speeches in those languages aligned with their interpretations. In 2005 the total number of tokens in the corpus was 172000 (Sandrelli – Bendazzoli 2005).

The depth of analyses performed on a corpus depends on the level of corpus mark-up. Raw corpora, i.e. running texts with no additional information in a form of tagging or parsing, may serve as a starting point for basic analyses: type/ token ratio, KWIC (keyword in context), frequency list or keyword list. Having a POS-tagged (Part of Speech) corpus at

one's disposal, it is also possible to establish its lexical density. The aforementioned methods of corpus analysis are discussed below.

Type/token ratio provides information regarding the overall relation between the number of different words (types) to the total number of tokens (sequences of characters delimited by spaces) in the corpus thus indicating how lexically varied a text is. It is counted by means of elimination of all duplicates and taking each word form into account only once. Such an analysis shall tell us that the "text fragment 'to be or not to be' contains six tokens, but only four types, as there are two tokens each of the type 'to' and 'be' " (Kenny 2001b: 34). Type-token ratio is dependent on the size of the corpus, therefore many concordancing applications produce average type-token ratio calculated as a mean of type-token ratios for every 1,000 words. This allows for comparison of corpora of unequal sizes.

Key word in contexts is a search method yielding concordance lines showing input called a keyword or node in its immediate context – a concordance line. It is possible to sort concordance lines e.g. alphabetically to the left or to the right of the node which makes it easier to identify recurring lexical patterns. Clicking on a selected concordance line usually presents the view of the keyword in the full context. In the case of a parallel corpus it is also possible to perform a parallel (bilingual) concordance search. The keyword is then presented on the screen in concordance lines in one language aligned with respective segments in the other language. Similar sorting procedures are usually available. The variety of options at one's disposal are, however, dependent on software.

Frequency list is a list of all words in the corpus with information regarding their frequency. The list may be sorted alphabetically or according to frequency. It is usually used to gain insight into vocabulary items best represented in a given corpus.

Keyword list makes it possible to determine which words are more characteristic in a given corpus with reference to another. The count involves a comparison of frequency lists of the two corpora in question, where one list is treated just as a reference. Software produces a list of words from list A (of a smaller, usually more specialised corpus), the frequency of which is untypically high with reference to list B (of a larger, usually more general corpus). The words are then ranked according to their 'keyness'.

The list of clusters, on the other hand provides information on words that most frequently co-occur in a corpus. Whereas, the function allows for search of clusters consisting of at least two words, it is searches for bigger clusters that offer most interesting insight.

Lexical density is a relation of lexical (content) words to function words in a corpus. The group of lexical verbs consists of nouns, main verbs, adjectives and adverbs. On the other hand, auxiliary verbs, modal verbs, pronouns, prepositions, determiners and conjunctions constitute the group of function words. It is possible to obtain such statistics if the corpus is POS-tagged, or alternatively by feeding a list of function words used in the analysed language into the corpus processing tool.

1.4. Universals as a major research area in Corpus-based Translation Studies

The idea to investigate translation universals developed in time and its most recent formulation has been shaped by three major factors: the distinctive nature of translated texts, the need of the Translation Studies to develop and establish themselves as a discipline and the influence of other closely related disciplines, such as linguistics.

Translation is a very specific form of communication using language that frequently attracts the attention of both lay people as well as translation scholars. While it is common for the former to find translations awkward or unnatural, translation scholars have for decades now made attempts to discover what makes translated language so distinctive. That it is different has been proved by the need to attach labels to it. The term *translationese* is often used by people other than translation scholars to refer to bad translations carried out by incompetent translators, whose language stands out as unnatural and full of errors. In Translation Studies, however, it acquires a slightly different, neutral meaning first mentioned by Toury (1980) and Gellerstam (1986) and referring to the nature of the translated language. The fact that translations are different from native texts was also noticed by Frawley (1984: 168-169), who viewed translation as a *third code* guided by its own rights, standards and structure, heavily influenced by the matrix information and the target parameters implying that translation is a special kind of code featuring the characteristics of two other codes: the source text and the target language.

The need of a research framework that could discover what contributes to the unique character of translated texts coincided with the stage of development of the Translation Studies that as an emerging discipline sought a rigid framework built on a set of rules, laws or principles. At that time Toury (1980) attempted to discover two major forces that determine the nature of translation and formulated two laws guiding translation: the

law of interference and the law of standardisation. The first assumes that certain characteristics of the source text are reflected in the target text and the second stipulates that the translators make an effort to standardise the target text in terms of language. In the years to come these two ideas tossed and turned by different researchers will also be transformed into two separate purported universals of translation.

In the consecutive decade Translation Studies experienced two important turns very much inspired by linguistics. The first involved the introduction of the concept of universals and the other the ‘import’ of corpus linguistics methodology to Translation Studies that offered a great potential in their investigation. When first proposed by Baker in 1993 the concept of universal features was already well established in linguistics, which certainly added to its appeal. The focus of research shifted slightly from laws guiding the process that shaped translated texts to the end product and its universals, defined by Baker as characteristic linguistic features

which typically occur in translated text rather than original utterances and are not the result of interference from specific linguistic systems... [but rather] a product of constraints which are inherent in the translation process itself, and this accounts for the fact that they are universal (Baker 1993:243-246).

Contrary to norms of translation proposed by Toury (1980), universals do not depend on either source or target language and are irrespective of social, cultural and historical context. Baker (1993) was also first to suggest that electronic corpora are an ideal solution to investigate linguistic features of translated language either in contrast with their originals or texts originally produced in the target language. She perceived corpus linguistics to be an appropriate tool to test the findings of research conducted by others on a small scale.

Included in her list were features such as a tendency towards explicitation (Blum-Kulka 1986; Toury 1991a), disambiguation and simplification (Blum-Kulka & Levenston 1983; Vanderauwera 1985), growing grammatical conventionality and a tendency to overrepresent typical features of the target language (Toury 1980; Vanderauwera 1985; Shlesinger 1991) as well as the feature of cleaning away repetitions from translations (Shlesinger 1991; Toury 1991b) (Kujamäki – Mauranen 2004: 1).

A number of scholars took up this idea and started to investigate the hypothesised translation universals, most notably explicitation, but also simplification and normalization or interference on large collections of machine readable texts. The outcomes of those studies are reported in greater detail in the following sections.

As this line of research gradually developed, it became clear that while seeking universals, translation scholars embarked on two different paths. This trend was first captured by Chesterman (2004: 8) who, in an attempt to introduce conceptual clarification put forward a division of translation universals into potential S-Universals and T-Universals: the former taking into account the ontology of translation and its relation with the source (S) text, the later focusing on specific features that distinguish translations from texts originally written in the target language. To the first category he includes, among other: the law of interference, the law of standardisation, explicitation hypothesis or reduction of repetitions. On the other hand, the features that typify translations when contrasted with texts originally produced in the target language encompass simplification, conventionalisation (in this thesis referred to as normalisation), untypical lexical patterning and under-representation of target language specific items. He stresses, however, that the two types of universals are to a certain extent interdependent e.g. interference might be a potential cause of untypical lexical patterning.

Chesterman's division seems to be very categorical, especially, that he stresses that it is impossible to investigate S-Universals without making reference to the source text, which implies that in terms of corpus methodology, they can only be analysed based on parallel corpora. Consequently, whatever feature is analysed on comparable corpora, it must be categorised as a purported T-Universal. In practice, however, much depends on how particular universals under investigation are operationalised in a given study and it turns out that a few universals categorised as S-Universals have been researched without reference to the source text, but based on monolingual, comparable corpora (see Section 1.4.2. and 1.4.5. for discussion on explicitation and interference).

It merits attention in that on the long way from laws to universals involving frequent shifts of focus from the process to the product of translation, many features purported to be universal have been operationalised in very different ways. This, on the one hand contributes to the deeper understanding of the intricacies of translation, but at the same time makes outcomes of several studies devoted to the same universal feature virtually incomparable. This lack of common platform of comparison and insufficient conceptual clarity is characteristic of the entire field of research devoted to universals, but particularly visible in studies devoted to explicitation (see Section 1.4.2.).

1.4.1. Simplification

Simplification may be observed at three different levels: lexical, syntactic and stylistic. According to Blum-Kulka and Levenston (1983: 119) lexical simplification is “the process and/or result of making do with less words”. Translators achieve lexical simplification by application of one of six solutions: use of superordinate terms, approximation, use of synonyms, transfer of SL word functions to TL equivalent, use of circumlocutions and paraphrases (Laviosa-Braithwaite 2001: 288). Simplification at syntactic level is demonstrated by rendition of complex syntactic structures with simplified ones (e.g. replacing non-finite clauses with finite ones). Stylistic simplification, on the other hand, involves breaking up long sentences, use of short collocations instead of elaborate ones, reducing repetitions and redundancies, shortening circumlocutions leaving out modifiers (Laviosa-Braithwaite 2001: 289). Referring to the phenomenon of simplification in general, Baker (1996: 176) claims further that the process is subconscious. Investigating translation strategies she also notes that in the case there is no equivalent, translators resort to superordinates. Similarly, in her analyses of fifty translations of Dutch literary texts Vanderauwera (1985) reports that sophisticated source language words were rendered with the use of simple, colloquial synonyms. Furthermore, she notices also stylistic simplification involving breaking up of longer sentences, omissions of redundant or repeated information, shortening of complex collocations etc.

A seminal corpus-based study of simplification was carried out by Laviosa (1998), who used English Comparable Corpus (ECC) a multi-source one million word monolingual comparable corpus of translated English (newspaper articles and prose) and texts originally produced in English. She discovered that irrespective of the source language the range of vocabulary used in translation corpus is narrower, which is indicated by the fact that:

- (a) translated texts have relatively lower percentage of content words versus grammatical words (i.e. their lexical density is lower);
- (b) the proportion of high frequency words versus lower frequency words is relatively higher in translated texts;
- (c) the list of a head of a corpus of translated text accounts for a larger area of the corpus (i.e. the most frequent words are repeated more often);
- (d) the list head of translated texts contains fewer lemmas (Laviosa 1998: 563).

She also found that translated texts had a lower average sentence length, which is a sign of stylistic simplification.

1.4.2. Explicitation and explicitness

Explicitation is probably the most unclear yet fiercely debated hypothesised translation universal. Scholars have looked at it from different angles not always having one definition or even one concept in mind. On the one hand, there are translation scholars who posit that explicitation is a phenomenon (referred to as a process, strategy or technique) that causes the translators to render information, which is implicit in the source text in an explicit way in the target text. On the other, there are supporters of the more contrastive approach towards translation who perceive it rather as a language specific tendency to encode information in a more or less overt or covert way. It appears that the concept of explicitation has originally emerged in Translation Studies as a hypothesised S-Universal, but scholars focusing on T-Universals decided to use the same term to denote what seems to be an increased tendency to explicitness. The confusion is aggravated by the fact that researchers rarely unambiguously state which definition they currently have in mind while investigating new instances of explicitation. Additionally, some of them use both terms interchangeably.

One of the first definitions of explicitation was provided by Vinay and Darbelnet (1995: 8) and referred to as “the process of introducing information into the target language which is present only implicitly in the source language, but which can be derived from the context or the situation”. This initial perception therefore points to the fact that as a result of certain process information that is implicit in the source language becomes explicit in translation.

The idea was further developed by Blum-Kulka in 1986 into explicitation hypothesis and has ever since been discussed and reshaped by a number of translation scholars. In its original wording it assumed that translations are more explicit than non-translations with respect to cohesive markers:

The process of interpretations performed by the translator on the source text might lead to a TL text which is more redundant than the source text. This redundancy can be expressed by a rise in the level of cohesive explicitness in the TL text. This argument may be stated as ‘the explicitation hypothesis’, which postulates an observed cohesive explicitness from SL to TL texts regardless of the increase traceable to differences between the two linguistic and textual systems involved. It follows that explicitation is viewed here as inherent in the process of translation (Blum-Kulka 1986:19).

As observed by Pym (2005: 2-3), while explicitation in the above definition is viewed as ‘the process of interpretations’ Blum-Kulka’s following analysis is restricted to linguistic qualities, in particular to markers of cohesive explicitness. In the end, one concludes that explicitation may be after all understood as redundancy i.e. “the unnecessary repetition of something that is already there [...] rather than a full process of something implicit [...] becoming explicit” Pym 2005: 2-3). Yet she does call it a process and not a technique or strategy. Furthermore she suggests that the process is inherent in translation, which has ever since been widely debated in the discipline giving rise to assumptions that it may be the sheer act of translating that causes explicitation (Klaudy 1996: 102-103).

In the following decades explicitation has been investigated in a number of studies, which approached it either from the linguistic angle or viewed as the process of making information implicit in the source text more explicit in the target text. Depending on the assumed perspective the phenomena held to be symptomatic of explicitation also differed.

Vanderauwera’s study published in 1985 presents observations, which are generally in line with Blum-Kulka’s but include a significantly larger variety of different phenomena as indicators of explicitation. She records that in the work of the translators of Dutch fiction explicitation can be observed in a number of applied techniques, such as:

the use of interjections to express more clearly the progression of characters’ thoughts or to accentuate a given interpretation; expansion of condensed passages; addition of modifiers, qualifiers, and conjunctions to achieve greater transparency; addition of extra information; insertion of explanations; repetition of previous details for the purpose of clarity; precise renderings of implicit or vague data; more accurate descriptions; naming of geographical locations and the disambiguation of pronouns with precise forms of identification (as reported in Laviosa-Braithwaite 2001: 289)

As showed above, the tendency to explicitation encompasses here an array of different phenomena that range from linguistic behaviours e.g. the use of interjections to strategies, which extend beyond the linguistic level such as providing more accurate descriptions or naming of geographical locations.

Also Sèguinot (1988: 108-109) chooses to adopt a broader definition than the one proposed by Blum-Kulka (1986) claiming that “explicitness does not only mean redundancy”. She claims that explicitation involves situations in which “something is expressed in the translation, which was not in the original” and “something which was implied or understood through presupposition in the source text is overtly expressed in the translation, or an element in the source text is given a greater importance in the translation

through focus, emphasis or lexical choice". In her study of French-English and English-French translations she observed increased explicitness achieved by improved topic-comment links, the addition of linking words and the raising of subordinate information into coordinate or principal structures. She does use both terms: *explicitness* and *explicitation* and while it seems that she understands explicitation as a process and explicitness as a textual feature she does not clearly state the differences between the two notions. This is misleading since it is difficult to ultimately determine, whether indeed she claims that the observed increased explicitness in translations has been an effect of the explicitation process.

Later Klaudy (1996: 102-103) makes a claim that explicitation may be motivated by four different factors and thus differentiates between obligatory explicitation, pragmatic explicitation, optional explicitation and translation-inherent explicitation. She argues that obligatory explicitation is enforced by grammatical and semantic differences between source and target languages and refers to the fact that translators have to make up for these differences to conform to the requirements of the target language. Pragmatic explicitation takes place when translators add context information that the source text recipients are aware of and the target audience lacks. Optional explicitation results from stylistic preferences of the translator or the target language community and aims to improve the presentation of the target text, but was it not applied the text would still be comprehensible. Finally, Klaudy proposes that there is a type of explicitation, which is translation inherent i.e. is caused by the nature of the translation process, but she does not provide examples and it is difficult to really comprehend this notion. She also does observe that translators not only tend to explicitate but also to implicitate i.e. render explicit information in an implicit way, but admits that the tendency to explicitation is stronger. In her distinction between different types of explicitation she does not refer at all to explicitness and so it is difficult to determine if she sees any difference between the two phenomena at all. Interestingly, one of the types mentioned by Klaudy i.e. optional explicitation does seem to correspond to the concept of explicitness, as it is perceived in contrastive approach to Translation Studies. The aforementioned approach looks at explicitness as overt encoding of information in a text, a feature that depends on the communicative conventions in different languages (e.g. House 1996, 2004). Kulka's optional explicitation must also refer to overt encoding, which results from stylistic preferences of the translator or the target language community i.e. the conventions appealing to either one or the other or both. Both

sources also claim that the phenomenon they define is only optional, since without them the message would still be fully comprehensible.

In the same year Baker (1996: 180) makes several proposals regarding the new methods of investigating universal features in a translation corpus. When suggesting a novel methodology to research explicitation with corpus linguistics means, she also briefly describes what she understands as explicitation:

I take “explicitation” to mean that there is an overall tendency to spell things out rather than leave them implicit in translation (Baker 1996: 180).

She points to the fact that this tendency can be observed in different textual phenomena such as text length (when compared with the source text, translation tends to be longer) and the use of explanatory vocabulary (she hypothesises that the translations make more explicit relations between propositions in text). While discussing *explicitation* Baker speaks of investigating the former on a parallel corpus and the later on a comparable corpus i.e. without referring to the source texts. This implies that her definition of *explicitation* does also include *explicitness* but she does not clearly state that.

Øverås (1998) investigated Blum-Kulka’s ‘explicitation hypothesis’ on an English-Norwegian parallel corpus of literary texts. She observed that translated texts display a greater level of cohesiveness, when compared to their source texts regardless of the directionality. Explicitating shifts, i.e. additions and specifications of lexical and grammatical items, were present in translations in both directions and outnumbered implicitating shifts in over 80% of texts. Furthermore, instances of explicitation at the level of lexis were more frequent than grammatical ones. Øverås suggests that explicitation may be a result of stylistic features prevailing in SL and TL, obligatory shifts pertaining to grammatical differences, culture bound translation norms and translation constraints.

Another corpus-based study of explicitation was carried out by Olohan and Baker (2000), who examined the frequency of occurrence of optional *that* with reporting verbs *say* and *tell* in translated and non-translated texts. They compared concordance results of two corpora: TEC (Translational English Corpus) and BNC (British National Corpus). According to quantitative results, optional *that* is far more frequently used in translated texts than in non-translated ones. Examining syntactic explicitation in the case of optional *that* Olohan and Baker (2000: 143) assumed that this form of explicitation is a sign of a subliminal process:

Unlike the conscious introduction of supplementary or explanatory material, a higher incidence of the optional *that* in translated English would provide evidence of inherent, subliminal processes of explicitation in translation. Translators clearly do not adopt a conscious strategy of spelling out optional syntactic elements such as *that* in reporting clauses more often than writers producing original texts in the same language (Olohan – Baker (2000: 143).

The study clearly shows that there is a difference between the number of occurrences of optional *that* in the corpus of translational and native English. There is, however, no accord among translation scholars as to what this fact really proves. Understanding explicitation in line with Baker's (1996) rather general definition, i.e. "the tendency to spell things out", one may indeed safely infer that translated texts manifest greater explicitation. For the advocates of explicitation viewed as a tendency to spell out things that were not spelled out in the source text Olohan and Baker's results do not bring anything new to the picture. As pointed out by Saldanha (2008: 22), "not all instances where the optional *that* has been spelled out in the translation are instances where there is implicitation in the source text". The examination of comparable corpora limits the analysis to the target text only and does not allow to investigate whether every case in which optional *that* has been used is an explicit rendition of implicit information. Becher (2011: 31) argues that "without a full answer to this question, Olohan and Baker's results cannot be interpreted properly, because the greater the number of source languages represented in the TEC stipulating a complementizer after reporting verbs, the more likely it becomes that the higher occurrence of reporting *that* in this corpus is the result of source language interference". So there are doubts as to whether Olohan and Baker's study provides a proof of explicitation. One can, however, interpret the increased frequency of optional *that* in translations as an indicator of a tendency towards overt encoding of information i.e. explicitness.

Olohan and Baker's was a large scale study that allowed to identify a global pattern, but the scope of analysis was limited solely to translations. To shed light on possible causes of such syntactic explicitness Kenny (2005) adapted their methodology to the requirements of a parallel corpus and examined source German expressions that were rendered with the reporting verb *say* followed by optional connective *that*. The study revealed that only a half of English explicit renditions (*say that*) were originally expressed with the use of German equivalent *dass*. Moreover, German originals using *dass* were mostly translated with *that*

rather than without it, which suggests that the overall tendency to explicitate was weaker in this particular case.

Martha Mutesayire (2004) examined lexical explicitation studying the use of apposition markers. Since apposition may be used to make the message clearer to the recipient, she believes that, in accordance with Klaudy's (1996) classification, it is a form of optional explicitation. She carefully examined various functions of apposition markers and selected a set that has an explicitation function. The set of reformulation phrases is further examined in the sub-corpus of TEC and sub-corpus of BNC on the example of reformulation phrases used in English (*e.g. that is, that is to say, to be (more) precise, to be (more) specific, namely*). She has discovered that apposition markers are significantly more frequent in translated texts than in non-translations. Mutesayire (2004: 54) claimed that such findings may be caused by the following factors: "the carrying over in the target text of specific features of the source text, (...) the low level of shared information between the translator and (...) the readership (...) and the translator's style".

Blum-Kulka's explicitation hypothesis and Baker's notion of explicitation has also been explored also by Papai (2004), who examined English-Hungarian translation with ARRABONA parallel corpus. She has found out that the effect of explicitation is achieved by translators with the use of three strategies: shifts in cohesion, addition of linguistic and extra-linguistic information and disambiguation. She has also noticed that the degree of explicitation in translated Hungarian texts is higher than in non-translated ones. Papai hypothesized that translated scientific texts may demonstrate a higher degree of explicitation than translated literary texts, but this hypothesis was not confirmed.

A strikingly different approach to explicitness in translation is offered by House (2004), who assumed that explicitness is a feature of discourse that varies across languages. She provides a qualitative analysis of English-German and German-English translations focusing on implicitating and explicitating shifts. The results of her examination point to unequal distribution of these shifts in respective directions: German translators explicitate more, while translations into English are more implicit. She provided examples of typically German pronominal adverbs not rendered in English translations and added in German ones. House (2004: 204) argued that explicitness and implicitness may be dependent on linguistic and non-linguistic factors. Translated texts may be more explicit or implicit due to translators' obligatory linguistic choices (necessary to produce a grammatically correct utterance) or optional linguistic choices (influenced by

communicative conventions of the target language). Apart from linguistic factors, the overall explicitness or implicitness can also depend on the translator's preferences, situation, or the requirements of the translation task.

At the same time Klaudy (2001) formulated her seminal asymmetry hypothesis that still drawn heavily on the notion of translation-inherent explicitation. It was extensively discussed in a joint article with Karoly (Klaudy – Karoly 2005). The asymmetry hypothesis stipulates that

explicitations in the L1→L2 direction are not always counterbalanced by implicitations in the L2→L1 direction because translators – if they have a choice – prefer to use operations involving explicitation, and often fail to perform optional implicitation (Klaudy – Károly 2005: 14).

The hypothesis has been tested qualitatively in the English-Hungarian and Hungarian-English translation of reporting verbs in literary texts. The study supports the hypothesis proving that in the analysed translations translators favoured more explicit forms and failed to perform implicitation. The authors argue, moreover, that if the tendency be confirmed on different language pairs, the asymmetry hypothesis may be a useful way to point to language-pair independent-universal feature of translator's behaviour.

Having critically reviewed key hypotheses on explicitation put forward in Translation Studies so far and not entirely satisfied by any, Pym (2005: 5) came up with his own model of explicitation placing it within risk-management framework. Understanding risk as a probability of an undesired outcome that could hamper cooperation between partners to the communication process, Pym claims that translators are likely to invest much effort in resolving high-risk problems and to resort to explicitation would be one possible way of handling them. His argument goes that for such reasons as 'prudence, Gricean cooperation, relevance to a new reception situation, the ethics of service (subservience), damage control or remedy' translators are more risk-averse. This risk-minimizing behaviour is further intensified by the fact that translation involves communication between contexts with few common references.

As the confusion kept growing more critical voices emerged in the discipline. Becher (2011) advocated the departure from the notion of translation inherent explicitation. In his own study he focused on identifying specific conditions in which explicating and implicitating shifts in translation are made assuming that "every instance of explicitation (and implicitation) can be explained as a result of lexicogrammatical and/or pragmatic

factors” (2011: 4). Having confirmed his initial predictions Becher specified a number of situations in which translators may feel motivated to explicitate or implicitate. He observed that translators tend to explicitate to save the readers from processing difficulties, to comply with target language communicative preferences, to minimise the risk of misunderstanding or due to lexical or syntactic mismatch between the source and target languages. On the other hand, they tend to implicitate on the following occasions: if a coreference relation explicit in the source text can be easily inferable, if there is no straightforward equivalent of a lexical item, “if the target language lacks a syntactic slot offered by the source language” or to comply with communicative preferences of target readers or improve stylistics and information structure (Becher 2011: 219).

As has been shown, the notion of explicitation has perplexed translation scholars for decades now and still there is no accord as to what explicitation really is and to what extent it is a translation inherent phenomenon. What seems to be common for all the abovementioned studies is that they all point to increased cohesiveness or overt encoding of information. To what extent it is a result of the translation process and not an effect of any contrastive differences between the languages or other factors it is hard to determine. What seems to stand out is that all translations appear to share increased explicitness, not necessarily caused by the explicitation process.

1.4.3. Normalisation

Another most frequently hypothesised universal involves normalisation, or “the tendency to conform to patterns and practices that are typical of the target language, even to the point of exaggerating them” (Baker 1996: 176-7). Similarly to other universals normalisation was analysed also without reference to computer corpora. Analysing translations of Dutch novels Vanderauwera (1985) found evidence of what she referred to as “lexical conventionality”, which manifests itself in shifts in punctuation, lexical choice, style, sentence structure, and textual organisation. Evidence includes adaptation of names and culture-specific items and reduced transfer of foreign language expressions, substitution of old-fashioned expressions by modern ones and translation of creative collocations with the use of common ones. Also punctuation is standardised, which shows in finishing

incomplete sentences, restoring quotation marks etc. As a result the translation reads better, is more idiomatic, coherent and familiar to the target reader (Vanderauwera 1985: 77-76).

Studies in normalisation at the level of punctuation were carried out by Malmkjær (1997), who analysed translations of Andersen's stories, and May (1997), who studied translations of Woolf's and Faulkner's novels. In all these cases it was found that the authors' unusual punctuation style was normalised.

In her aforementioned study, Øverås (1998) noticed additionally that unusual collocations and metaphors used in English-Norwegian translations of literary texts are also subject to normalisation (which she calls neutralisation).

Scott (1998) analysed the translation of a Brazilian Portuguese novel written in a peculiar, difficult and complex style looking at normalisation. Scott used WordSmith Tools to trace textual changes in the English translation and establish how they contribute to the overall effect of normalisation. On examination of the text Scott distinguished two poles of normalisation; one resulting from the TL systemic constraints and the other caused by translator's preferences. Scott "suggests that the translator's choices, conscious or not, obligatory or optional, cause the breaking up of cumulative effect" (after Laviosa 2002: 69).

Lexical creativity and lexical normalisation constitute the focal point of Kenny's (1999) study conducted on a two-million word parallel corpus of German literary texts and their English translations. Kenny found out that 44% of creative forms (creative orthography, derivation, compounds) were normalised in translation and that compensation was present in the case of only 16% of normalised hapax legomena (word forms that appear only once in the corpus) and absent in normalised collocations. Kenny claims that normalisation may be influenced by the information about the author, the translator and the publisher.

Baker (Baker 2004, as reported by Olohan 2004: 151) tests translator's tendency to employ strategies which would make their work meet their reader's expectations. She puts forward a hypothesis that a higher number of semi-fixed and fixed expressions occur in translations than in non-translations. Normalisation is the tendency to conform to patterns "typical of the target language, even to the point of exaggerating them" (Baker 1996: 176-7). Thus it may be argued that by attempting to meet the readers' linguistic expectations the translators reveal a tendency to normalisation. It seems therefore that although Baker (2004) does not explicitly state that, her findings are also a proof of

normalisation. She has chosen popular fixed and semi-fixed expressions such as *on the other hand, at the same time, once and for all, in other words, in a manner of speaking, that is, that is to say, once and for all, when it comes to, I thought to myself* and checked their recurrence in TEC (Translational English Corpus) and the BNC (British National Corpus). She found that these expressions were more frequently used in translational English than in non-translations. Furthermore, it appeared that some of the phrases were favoured by one translator, whereas other expressions by others.

1.4.4. Levelling out

Levelling out or convergence is commonly understood as the tendency of translated texts to be similar to one another with regard to particular linguistic features. Baker (1996:184) uses the term levelling out¹, while Laviosa (2002: 73) calls it convergence.

The hypothesis that translated texts share certain linguistic features was supported by Laviosa's (2002: 73) study comparing lexical density and type/token ratio as well as sentence length in translational and non-translational corpora. She found out that translated newspaper articles are more uniform with respect to these parameters than comparable non-translated articles.

1.4.5. Interference

Interference is a puzzling phenomenon that has been observed in various language contact situations and as such it is also prone to occur in translation. The concept first emerged in the field of second language acquisition and referred to deviations in norms of a language that can be observed in speeches produced by bilinguals "as a result of their familiarity with more than one language" (Wenreich 1953: 1). Since translation involves bilingual processing Mauranen (2004: 67) claims it is only reasonable to assume that interference also occurs in translation.

¹ Levelling out is understood by Baker (1996: 176-177) as "the tendency of translated text to gravitate around the centre of continuum, rather than move towards the fringes". She borrowed the term from Shlesinger (1989: 170), who coined it to depict shifts from orality to literateness and vice-versa in simultaneous interpretation.

The status of interference in Translation Studies is, however, not clear in that it is frequently conflated with transfer, it is either believed to be a more generic, neutral term or contrasted with positive transfer. Also its status as a universal feature of translation has not been acknowledged by all.

In his formulation of the Law of interference Toury (1995: 275) stipulates that “in translation, phenomena pertaining to the make-up of the source text tend to be transferred to the target text”. What follows is an explanation that the source text phenomena may be imported in the form of negative or positive transfer, but the tendency to interference is in a way default and that extra effort must be made by the translator to produce interference-free output. While Toury perceived positive and negative transfer as a form of interference, Mauranen (2004: 71) reports that transfer and interference are used either interchangeably or to denote opposite notions. In the later variant transfer is viewed as positive or neutral and interference as a negative phenomenon. Positive or neutral transfer results in renditions that bear the characteristics of the source text/language, but also are in compliance with target language norms. Interference (or negative transfer), on the other hand, leads to formulations that do not conform to the target language conventions and stand out as awkward and foreign.

While Toury views interference as a default element of translation, Baker (1993: 243) argues that a feature of translation may be deemed a translation universal if it is not a result of interference. Mauranen (2004: 79) tries to reconcile both views by claiming that the source language does shape the target text, but it cannot be the only influencing factor, since translations from different source languages still resemble each other. Thus she suggests that

interference (or transfer) is best conceptualized as one of the universal tendencies, on a high level of abstraction, precisely on account of predictably taking place in each language pair involved in translation (Mauranen 2004: 79).

She arrived at this conclusion having analysed the scope of deviations of comparable corpora of translational Finnish and native Finnish. The examination focused on differences in frequency ranks between a corpus of native Finnish, a corpus of mixed translational Finnish (i.e. combined translations from Russian and English), a corpus of Finnish translations from Russian and a corpus of Finnish translations from English. The results point to a greater similarity among translational corpora than between them and

native Finnish texts. Yet a clear source language effect could be observed since corpora of translations from English deviated less from native texts, while translations from Russians resembled native Finnish more.

1.5. Translation universals reflected in selected interpreting research

Pym (2007: 2) notices that although conference interpreting is one of the modes of translation, it is rarely mentioned in literature on universals. The deficiency of research in this domain may be most probably attributed to the difficulties in obtaining interpreting data. Additionally, be it authentic data or experimental recordings they both require laborious transcription before the actual analysis is possible. Yet, there is a great potential in the corpus-based methodology developed for the investigation of translation, which should be exploited more extensively by interpreting scholars. Such an approach offers an opportunity to compare and contrast interpreting with translation and extract common features characteristic of both processes.

Meanwhile the major focus of interpreting researchers seems to be on investigating the same phenomena with different methods, more adjusted to the nature of interpreting and fitting the requirements of small scale research mostly. As such, this prevalent trend offers insight of incredible value, but is not directly comparable to the outcomes of translation research.

1.5.1. Simplification

Probably the only study that has so far adapted corpus-based methodology used in Translation Studies is devoted to simplification. Laviosa's (1998) methodology for the investigation of simplification as a universal feature of translated texts based on lexical density and list head analyses was applied by Sandrelli and Bendazzoli (2005), who tested the relevance of simplification hypothesis for simultaneous interpreting. The study focused on the examination of sub-corpora of EPIC (European Parliament Interpreting Corpus) i.e. subcorpora of interpreted English, subcorpora of interpreted Italian and subcorpora of speeches originally produced in those languages. The findings were, however, not exactly

in line with Laviosa's (1998) observations. First, it seems that lexical density in simultaneously interpreted texts is only slightly different to the one of speeches originally produced in English and Italian. Sandrelli and Bendazzoli (2005) attribute it to the peculiar character of interpreting, namely:

...the specific text production conditions, i.e. the pace of the incoming speech is imposed by the source speaker and the interpreter has to assemble the target speech practically "on-line", chunk by chunk, by selecting and re-arranging information to suit the norms of the target language. 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 and Bendazzoli 2005: 15).

Lexical density in the case of interpretations from Spanish into Italian i.e. the only combination of two Romance languages, proved to be an exception and confirm Laviosa's observations. Second, analysis of list head of the subcorpora seems to give contradictory findings. Whereas in the case of original Italian speeches and interpreted Italian, the results are completely different from Laviosa's observations on translational English, the analysis of interpreted English and original English proves that, similarly to the case of translational English, "the nuclei of words most frequently used in speeches interpreted into English are less varied and account for a larger part of the corresponding sub-corpora" (Sandrelli – Bendazzoli 2005). The analysis shows yet another interesting fact, this time related to the distribution of lexical and function words in the list heads. The list head of the corpus of speeches originally produced in English consists of a lower percentage of lexical words than both interpreted English sub-corpora (it-en and es-en). The researchers believe that this may be attributed to the interpreter's reformulations, self corrections or instances of expansion and explanation of the source text. Such instances may have increased the number of lexical words in the corpus. Interestingly the latter observation is confirmed only in the subcorpus of speeches interpreted from English to Italian, whereas a similar examination carried out on the subcorpus of speeches interpreted from Spanish lead to the observation of an opposite trend.

The unquestionable advantage of Sandrelli and Bendazzoli's study is that the authors strive to replicate Laviosa's methodology, which makes the parameters comparable and thus facilitates the comparison of patterns characteristic of translation and interpretation. Additionally, EPIC is parallel and bi-directional thus offering the unique opportunity to consult the source speeches of interpretation and to observe if the patterns

discernible in one language pair are also revealed in a different pair. On the other hand, it appears that the make-up of EPIC may question the validity of the results obtained by Sandrelli and Bendazzoli. First, the reference corpus of speeches originally produced in English is compiled of speeches delivered by both native and non-native speakers. The idea behind reference corpora used in Translation Studies to help shed light on the recurring patterns of translated and non-translated language is that they consist of text produced by native speakers. Including other texts may introduce into the corpus patterns that are not really typical of the investigated language and skew the outcome of analysis. Second, the corpus of interpretations into English contains also texts interpreted by one non-native speaker of English carried out in *retour* mode (from one's mother tongue to an active foreign language), which most likely is characterised by different patterns than interpretation into mother tongue. Thus both the interpreting corpus and the reference corpus of English in EPIC are tainted by features of non-native language, which makes it difficult to establish which of the features could be accounted for by the interpreting process and which by the non-native use of language. Moreover, the parameters used in the investigation of simplification, such as the calculation of list head proportion in the corpus are very susceptible to corpus size and the compared corpora are very uneven: 42705 tokens in Org-en, 6708 in Int-it-en and 12995 in Int-es-en. In conclusion, Sandrelli and Bendazzoli's study on EPIC is a major advancement in Corpus-based Interpreting Studies, but it is not free from some flaws.

1.5.2. Explicitation

In her MA thesis Shlesinger (1989) tested Blum-Kulka's explicitation hypothesis, which however, was not supported by the findings. Reviewing Shlesinger's analysis² Pym (2007: 11-12) suggests that lack of explicitation in simultaneous interpreting is not surprising and may be attributed to risk-reducing strategies. He claims that the speaker and the audience possess a greater knowledge on the subject matter than the interpreter. In the case the interpreters are not sure they would rather use superordinates and resort to implicitation in their rendition.

² Shlesinger's (1989) MA dissertation has not been published and is not generally available for reference.

[B]ecause she lacks knowledge of the context, the interpreter actually misses some of the cohesion patterns, therefore resulting in less explicitation in the rendition. When you are not sure of what is going on, you cannot risk underlining relations that are no more than guesswork. A far better strategy in such situations is to say less, to use superordinates in cases of doubt, and to stay close to the given cues, since even if you don't understand, there is a good chance the audience will. Hence the use of considerable lexical implicitation (Pym 2007: 11-12).

Pym explains further that explicitation in translation was supposed to make up for the context the recipients lack. In the case of simultaneous interpreting it is usually the speaker and the recipients who have more background knowledge on the subject matter. In the case of highly specialised conferences the risk of misunderstanding is extreme. In such cases “explicitation is too risk-laden, and so one logically turns to implicitation” (Pym 2007:12).

A few years later Shlesinger (1995) investigated the rendition of cohesive ties in interpreting. She compiled a collection of interpretations of the same English text delivered in simultaneous mode by thirteen advanced students of interpreting. Shlesinger notes that interpreters' renditions are subject to three major constraints regarding time, linearity and knowledge. First, the ability to render the input is dependent on the time that can be allocated for hearing and processing it. Secondly, the interpreter has no access to the entire text at once, which means that a wrong application of global and local strategies may lead to misunderstanding. Finally, there are bound to occur discrepancies between the knowledge that the speaker expects from the audience and the one possessed by the interpreter. The difficulty lies in the topic of the speech, as well as the fact that many impromptu speeches seem to be difficult to comprehend also to the target audience and pose exceptional perceptual difficulties.

In the light of the constraints reiterated above it does not seem surprising that the majority of shifts observed in the study involved complete omissions and cuttings in all examined categories of cohesive ties. The omissions were most frequent in case of elements bound to be known by the audience, requiring low-frequency vocabulary or referring to non-essential information. Many omissions regarded elements in sentence-final positions, discourse markers or certain categories of conjunctions (dismissive, temporal and emphatic).

Shlesinger's findings (1995: 211-213) again suggest that interpreters tend mostly to use omission. She has, however, found that one type of cohesive ties i.e. substitution (including ellipsis) was mostly replaced with “some form of lexical cohesion” and marked

that such an outcome may also be seen as lending further support to the premise that the explicitation hypothesis (Blum-Kulka 1986) applies to oral as well as written translation.

To unveil the causes of the hypothesised greater explicitness of interpreted texts Gumul (2006) examined explicitation in simultaneous interpretation qualitatively by comparing source and target texts and analyzing interpreter's comments. She discovered that explicitation in simultaneous interpreting is in the majority of cases a subconscious process, which she attributes to time pressure and a 'piecemeal picture of text structure' (2006: 184) and that most of the explicating shifts in interpreting took the form of cohesive explicitness i.e. additional connectives, reiteration of lexical items, shifts from reiteration in the form of paraphrase into identical or partial repetition, shifts from referential to lexical cohesion and filling out elliptical constructions.

1.5.3. Normalisation

The tendency to normalisation has not been extensively explored in the field of translation and even less in interpreting. The only observations regarding normalisation in interpreting come from a study, where normalisation was not the major focus of research. Analysing consecutive interpreting from Hebrew into English provided by courtroom interpreters Shlesinger (1991) found evidence of normalisation. Interpreters completed unfinished sentences, provided grammatical rendition of ungrammatical source utterance, omitted false starts and self-corrections.

1.5.4. Levelling-out or convergence

Shlesinger (1989) examined whether simultaneous interpretation had an effect on orality or literateness of interpreted texts. What she found was that "oral texts take on more literate features in simultaneous interpreting and literate texts become more oral. In other words, the process of translation tends to move texts more towards the centre of the oral-literate continuum, to locate them away from either extreme" (Baker 1996: 184). Shlesinger refers to this phenomenon as equalizing, whereas Baker calls it levelling out.

In her later study Shlesinger (2009: 244) compiled corpora of six translated (8,968 words) and six interpreted (8,317 words) outputs in Hebrew. The target texts consisted of oral and written renditions of one English text provided by six professional interpreters/translators. The texts were interpreted by the subjects more than three years before they completed the translation task, hence “the participants were unlikely to recall either the text itself or the strategies they had used to render it... [and] the two sets of texts may be seen as independently produced outputs based on the same input” (Shlesinger 2009: 239-240). The author classifies her corpus as *comparable intermodal* i.e. one which consists solely of translations in different modalities or modes. She managed to extrapolate a set of characteristic stylistic and pragmatic features distributed unevenly between interpreted and translated output. The analysis of lexical variety by means of type-token ratio (TTR) proved that values were higher for the translation corpus as a whole and for each of the six individual texts. Type-token ratio of the interpreting corpus (as a whole) equaled 0.655, whereas the translation one 0.735. Examination of verb patterns proved that simple verb patterns are preferred in simultaneous interpreting, whereas morpho-syntactically complex patterns in translations. Yet another finding concerns the number of definite articles in both corpora: it was considerably more common in the oral mode of translation. Shlesinger (2009: 247) found also that verb-based constructions, which are more accessible and more dynamic occur more frequently in interpreting corpus than in translations.

1.6. Corpus-based investigations of translator’s style

Research on translator’s style is a relatively new trend in Corpus-based Translation Studies. Translators’ work has for long been by many perceived as, or expected to be, “invisible” i.e. leaving no traces. Contrary to that presumption Hermans (1996: 9) developed the notion of translator’s presence and clearly acknowledged that “that other voice [i.e. the translator’s] is there in the text itself, in every word of it”, claiming, however, that having translated the text the translator “disappears without textual trace” and his voice, although occasionally present in paratextual notes “may remain entirely hidden behind that of the Narrator, rendering it impossible to detect in the translated text”.

Baker (2000: 245) argued that new advances in corpus linguistics may foster investigating translator's presence in translation, and in particular the style of a literary translator. For the purpose of her study Baker proposed a new definition of style, which she understood as

... a kind of thumbprint that is expressed in a range of linguistic – as well as non-linguistic features. [...] style might include the literary translator's choice of material to translate, where applicable, and [...] consistent use of specific strategies, including the use of prefaces or afterwords, footnotes, glossing in the body of the text, etc. [it] must focus on the manner of expression that is typical of a translator [...], attempt to capture the translator's use of language, his or her individual profile of linguistic habits, compared to other translators. [Style] is a matter of patterning: it involves describing preferred or recurring patterns of linguistic behaviour... (Baker 2000: 245).

Therefore, what Baker (2000) is mostly looking for is subtle, unobtrusive linguistic habits, patterns of choice, which are predominantly subconscious.

She applied this novel perception of style in the comparison of works by two literary translators: Peter Bush and Peter Clark focusing mostly on the level of repetition of the respective translations and predominant verb patterns.

The first comparison is made on the basis of type-token ratio indicating the level of repetition in the analysed texts. Baker's (2000: 250) analysis proves that type-token ratio is higher in the case of Peter Bush's translations, which shows greater lexical variation. The analysis of individual translations with respect to type-token ratio demonstrates that the collection of texts translated by Bush is less homogenous, whereas Clark's translations are similar. Clark favours shorter sentences than Bush, which Baker interprets as an attempt to mediate Arabic texts and make them less challenging in terms of linguistic features. Baker (2000: 251) analyses also how both translators use the verb *say* (in all forms *say, says, said, saying*) which happens to be the most frequent reporting verb in English. She found that Peter Clark makes a heavy use of the verb, especially favouring past tense and tends to use other reporting verbs also in past tense, rarely resorting to third person present forms. Comparing fragments of the original text Baker (2000: 252) found that past tense is sometimes his own choice independent of the source text. Baker concludes that such choices influence overall level of formality of the text and a sense of immediacy. Clark also prefers direct speech over indirect speech, which may be perceived as an attempt to make the utterances more directly available to the reader and make it more accessible. Bush, on the other hand uses indirect speech, which poses no clear cut boundaries between the end of

quotations and the narrator thus engaging the reader more into the fictional world (Laviosa 2002: 84). Baker (2000: 254) notices furthermore that “there seems to be a strong preference for modifying verbs of speech, for adding something about the manner something was said” in Clark’s texts e.g. *angrily, apologetically, decisively, with a slight foreign accent*.

Baker’s innovative approach to translator’s style gave rise to a number of studies investigating recurring linguistic and non-linguistic patterns in works of individual translators. The idea was also employed by Olohan (2004: 160-166), who examines if translation strategies advocated by Venuti are indeed reflected in his translation. She investigates his lexical choice using keywords analysis. Keywords analysis is usually used to compare a specialised corpus with a general one. Olohan (2004), however, investigates stylistic choices of the translator. Venuti (1995: 1) believes that translators and publishers tend to attach too much importance to the issue of fluency and idiomaticity, which makes the translator less visible. In 1998 he sets out to combat this trend in his translation by using specific strategies to produce a foreignising effect and make the readers more aware that they are reading a translation. Choosing a collection of Buzzati’s short stories and Tarchetti’s novel (*Passion*) translated by Venuti, Olohan (2004) aims to examine, whether keywords analysis could provide insight into his foreignising strategies.

It appears that Venuti often resorts to *yet, nonetheless, nor* and *thus* typical of academic discourse. He also opts for the use of *wherein*, which is often considered archaic, in the translation of *Passion*, and decides for *in which* translating Tarchetti’s works. Olohan (2004: 164) suggests that Venuti’s choices are deliberate “as he chose not to use *wherein* at all in the translations of more contemporary writer Dino Buzzati”. Further characteristic lexical choices observed by Olohan include deliberate choice of *toward* or *towards* in different texts, excessive use of *shall* and *that*. The latter may be a sign of heavy use of optional *that*, which is not only a sign of explicitation, but also contributes to the potentially unwieldy language. The keyword list analysis shows also which adjectives and verbs are most favoured by Venuti. Olohan (2004: 166) then proceeds to compare the translator’s own comments on the strategies he applied. Venuti (1998: 14) admits that he deliberately favoured archaisms over current forms to indicate “the temporal remoteness of the Italian texts, their emergence in a different cultural situation at a different historical moment” e.g. by choosing *nor could I ever* over *and I could never*. Olohan (2004: 166) observes that this may explain the high number of *nor* clauses in his corpus. Furthermore, Venuti preferred

formal verbs to colloquial ones, which is reflected in the keyword list, where predominant verbs are *resume*, *seize*, *render*, *remain*, *possess*, *abandon*, *depart*, *explain*. Another strategy involved the use of Briticisms in the translation of *Passion* aimed primarily at American readers, which shows in the use of *towards* (British) rather than *toward* (American). Finally, in certain cases Venuti preferred to use Italian expressions instead of English ones. This phenomenon is evidenced at lower positions of the keywords list occupied by *adieu* (190th position), *amour* (261st position) or *propre* (435th).

Thus Olohan successfully proves with the use of the analysis of a keyword list that linguistic choices taken by Venuti in the process of translation do contribute to the foreignising effect.

The analyses of translator's style reported above introduced the topic to the field of Corpus-based Translation Studies, but focused rather on what corpus tools had to offer in investigations of the kind and did not provide any broader theoretical framework for such research. In her recent study, Saldanha (2011:31) examines different definitions of writer's and translator's style in an attempt to fill this gap and put forward a clear theoretical and methodological framework for the investigation of translator's style. She proposes the following revised definition of translator's style:

A 'way of translation' which:

- is felt to be recognizable across a range of translations by the same translator,
- distinguishes the translator's work from that of others,
- constitutes a coherent pattern of choice,
- is 'motivated', in the sense that it has a discernable function or functions, and
- cannot be explained purely with reference to the author or source-text style, or as a result of linguistic constraints (Saldanha 2011: 31).

She then sets out to test her definition by comparing in a corpus-driven investigation the style of two different translators: Peter Bush and Margaret Jull Costa. Having compiled two corpora containing different works of both translators, Saldanha investigates the use of italics as a means of highlighting foreign words, emphatic italics to accentuate marked information and use of *that* connective after reporting verbs *say* and *tell*. She found that translations by Peter Bush show higher incidence of source culture borrowings marked with italics in the target texts, lower frequency of emphatic italics and does not favour *that*

connective after reporting verbs *say* and *tell*. Margaret Jull Costa, on the other hands, does not often resort to borrowings, but shows a greater tendency to mark information with italics, even if that strategy has not been used in the source texts. She also is more prone to use *that* connective after the investigated reporting verbs. It may be thus concluded that both translators differ from each other with respect to these features. Saldanha, compared these findings also against a set of larger translational and native English corpora to see how they correspond to global patterns. It was observed that the outcome of the idiosyncratic analysis deviate from the norm e.g. while the incidence of optional *that* is far lower in translations produced by Peter Bush than in the Translational English Corpus, the frequency of this phenomena in Margaret Jull Costa's translations is higher. Thus Saldanha proved that the examined features allow to discern distinctive styles of the two translators. She argues that the choices made by individual translators were motivated by different global strategies. The texts produced by Peter Bush aim at stressing cultural differences and do not prioritise smooth reading hence a more frequent use of borrowings in italics and low frequency of optional *that*. Margaret Jull Costa, on the other hand, favours greater familiarity and more explicit renditions.

1.7. Conclusions

This Chapter has provided a brief overview of the development and the major trends in corpus-based research on universals and style in translation and their reflection in Interpreting Studies.

It proves that corpus methodology turned to be a valuable advancement in translation and interpreting research allowing for quantitative verification of hypotheses that have been developed over decades of case studies based on small volumes of texts only. Corpus linguistics has been confirmed as a useful tool allowing to spot recurring patterns of translated language in macro analyses of large compilations of translations from different languages. It has also been successfully applied in micro analyses of features that typify the style of individual translators carried out on much smaller collections of translations produced by a single author.

What is also unveiled is the acute imbalance between Corpus-based translation and Interpreting Studies. Corpus-based research on interpreting universals is in its infancy,

while interpreting style has been virtually unexplored. So far, most trends hailed to be universal of translation have not yet been investigated on interpreted discourse and it yet remains to be seen how much these two modes actually have in common and what makes them different.

These lacunae should be further filled with studies like Sandrelli – Bendazzoli’s (2005) research of simplification in simultaneous interpreting in EPIC, but need also to be complemented by analyses of individual idiosyncratic tendencies to prove if the investigated patterns are indeed truly universal or merely prevalent. Such patterns should be examined on texts produced in different conditions, which would allow to eventually identify factors that call them into existence.

It is hoped that the present dissertation will help bridge the gap between Corpus-based Translation and Interpreting Studies by applying the same yardstick that has already been used in the exploration of the translated language to the one of interpretation, thus making the results directly comparable. It will also attempt to combine the methods developed for the investigation of universals with the ones adopted in the examinations of style to analyse both notions more comprehensively.

Chapter 2: Objectives and methods

2.1. Objectives

As has been shown in the previous Chapter, while translation universals have frequently been investigated on translation corpora by translation scholars, little has been done to verify if the universals are also prevalent in interpreting. Universals in this mode have not been so extensively researched, most probably due to the fact that such studies are difficult to carry out on a large scale. Authentic data are not easily accessible and compilation of a corpus requires meticulous and time consuming transcription work.

It is hoped that the understanding of interpreting universals and interpreting style will be increased by the present corpus-based investigation. The overriding goal of this study is to verify whether three major features believed to be translation universals i.e. simplification, explicitness and normalisation are at all valid in the case of simultaneous interpreting and to what extent they are affected by idiosyncratic stylistic preferences. To examine the potential ‘universality’ of these features the analysis will involve interpretations into one target language (English) from four source languages (French, Spanish, German and Dutch). The study will also investigate whether there exist any potential differences in putative universal features present in texts interpreted into English from Romance languages (French, Spanish) and Germanic languages (German, Dutch).

The present analysis is also set to determine, whether interpreting universals are equally reflected in the performance of individual simultaneous interpreters or, whether

different realisation of these features would rather lead to identification of a particular interpreting style of a simultaneous interpreter.

It will also be examined if distribution of those potentially universal features of interpreting differs in interpreter's performance when the source language changes. In other words, it is expected that the study helps to answer the question whether the way individual interpreters perform interpretation is in any way affected by the source language. Different prominence of the investigated features in one's interpretations would be considered characteristic of individual interpreting style.

It might be the case that one's speaking style affects one's interpreting performance. For that reason the present analysis will also examine non-interpreted discourse produced by the interpreters and focus on a comparison of interpreted and non-interpreted performance with respect to potentially universal features of interpreting.

Thus the study will attempt to provide answers to the following research questions:

- (1) Are simplification, explicitness and normalisation interpreting universals?
- (2) Are individual interpreters consistent in their tendencies to simplify, explicitise and normalise regardless of the source language?
- (3) Do individual interpreters differ from each other in their tendencies to simplification, explicitness and normalisation?
- (4) Is an individual's interpreting style similar to their speaking style?

Due to a number of constraints inherent in the process of simultaneous interpretation, most notably time constraint and linearity constraint, as well as the different mode of delivery of simultaneous interpretation as opposed to translation, it is difficult to anticipate any particular outcomes. The study will, however, start from the following preliminary hypotheses:

- (1) All simultaneous interpretations tend to be more simplified, more explicit and use more normalised language than non-interpreted discourse.
- (2) Individual interpreters are consistent in their tendencies to simplify, explicitise and normalise regardless of the source language.
- (3) Individual interpreters differ from each other in their tendencies to simplification, explicitness and normalisation.

(4) An individual's interpreting style is similar to their speaking style.

The first of the above hypotheses will be tested in the macro analysis presented in Chapter Three, while the next three in the micro analysis depicted in Chapter Four. They will all be validated in line with the methods and on the set of data described in the following sections.

2.1.1. Operationalisation of research questions

The overriding aim of the present study is to verify if three patterns recurrent in translation, namely, the tendencies to simplification, explicitness and normalisation are interpreting universals and whether their prominence constitutes a feature of a distinguishable interpreting style. It follows that the present study will investigate the three features at a macro level of a large number of interpreted texts produced by different interpreters working from different languages and at a micro level of a smaller number of interpreted texts produced by individual interpreters working also from different languages. The notion of style and the notion of universality are very general, evoke very different connotation and their exact understanding within the present study must be clearly specified.

The term universality is in the present study perceived in a very categorical way. A feature or a pattern could only be considered an interpreting universal if it does typify all investigated interpretations, and whenever contrasted with native speeches, interpretations would display a greater prominence of that particular feature. It follows that for a feature to be regarded universal, it must be proved to be independent from interference. Consequently, should the tendency to simplification, explicitness or normalisation be manifested very differently (when compared to native speeches) in interpretations from different languages, it would not be accepted as an interpreting universal. It is also believed that if any of the features does indeed prove to be independent from the influence of all of the source languages investigated here, then it may only be regarded as a likely candidate for an interpreting universal, as it should be also tested on different language pairs in the future. Once, however, the analysis shows that one of the features is not to be traced in an interpreting corpus or interpretations of an individual, it cannot be regarded universal.

Interpreting style, on the other hand, following the path set by Baker (2000: 45) and Saldanha (2011: 31) in particular, is viewed here as a way of interpreting manifested by a consistent prominence of one or many linguistic features in different interpretations that makes one's interpretations distinguishable from others, as well as from their own speaking style.

While at the first glance the two concepts may seem contradictory, as one may doubt if a feature that is universal can at the same time be characteristic of an individual. Universality and style do not have to be mutually exclusive. It seems perfectly viable that compared to native speeches, all interpretations could be always more explicit and one interpreter could be consistently significantly more explicit than others in all his or her interpretations, regardless of the source language or the original speaker. It is by no means argued, however, that interpreting style may be manifested only by one's tendency to simplification, explicitness or normalisation, but it is believed that certain idiosyncratic differences may be revealed by some or all of the investigated features.

The features in question, that is simplification, explicitness and normalisation have been so far approached from different angles and the perspective assumed in the present study must be clarified.

In the investigation of simplification this research follows the approach presented by Laviosa (1998: 563). The focus here is placed on lexical simplification demonstrated by texts' tendency to greater repetitiveness, lower level of informativeness and lower lexical sophistication. In other words, when two texts are compared, the more simplified would be the one where more vocabulary items repeat, which contains more function words (and fewer content words) and is made up in a greater proportion of the most frequent English words. In corpus methodology those features translate respectively into the following parameters : a greater proportion that corpus list head represents in the entire corpus, lower lexical density and higher incidence of high frequency words (for more details see Section 2.3.1.).

Due to a significant inconsistency and confusion in the field of Translation Studies around the concept of explicitation and explicitness (see Section 1.4.2.), it should be stated especially clearly what exactly is understood as the tendency of a language to be more explicit. The present study focuses on the surface strata of language and therefore investigates explicitness understood as the tendency to encode information in an overt rather than covert way. It is assumed here that increased explicitness may, but does not

have to, be a result of the process of explicitation, in which the information implicit in the source text becomes explicit in the target text. Thus the process of explicitation, as such, remains out of the scope of investigation of the present study. The tendency to explicitness is operationalised as increased use of optional complementizer *that* after reporting verbs, higher frequency of linking adverbials and greater incidence of apposition markers (see Section 2.3.2.). The three aforementioned features are not compulsory elements of language, hence the tendency to spell them out is here perceived as greater explicitness. It merits attention that the three parameters so far in the field of corpus based Translation Studies have been mostly used to testify for or against explicitation (see Section 1.4.2.). In the light of the recent debate on the nature of explicitation, the author of the present research decided to take a stance that the process of explicitation cannot be investigated without the source texts. As the source texts are not investigated within this study, the only possible tendency that could be observed is explicitness.

Finally, the third examined feature is normalisation involving a more conventional and conservative use of language manifested in more excessive use of standard, most frequent patterns. It is believed that a text characterised by the tendency to normalisation makes more frequent use of elements that may be deemed standard in a given register such as lexical bundles or fixed expressions. The latter lexical features are used to investigate normalisation in selected corpora (see Section 2.3.3.).

The present study consists of two analyses: macro analysis of interpreting universals and micro analysis of interpreting style.

The aim of the macro analysis is to establish whether simplification, explicitness and normalisation could be universal features of interpreting i.e. if each of the three features are significantly more pronounced in interpretations into English from four different languages: French, Spanish, German and Dutch, than in native English speeches.

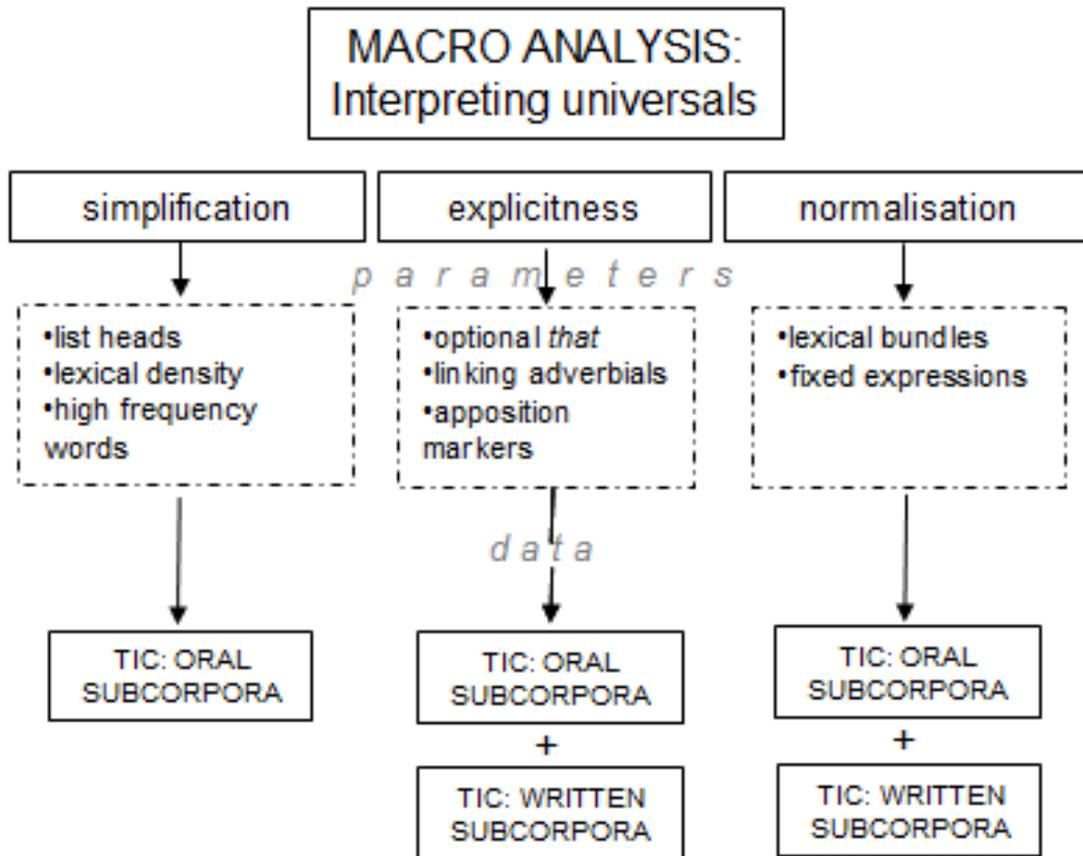


Figure 1. The structure of macro analysis of interpreting universals depicting the investigated features, parameters and data sets.

As presented in Figure 1 each purported universal of interpreting will be examined on a set of comparable corpora comprising transcriptions of interpretations into English from the four languages enumerated above (see Section 2.2.1. below). The results obtained from interpretation corpora will be immediately contrasted with the ones of corresponding translation corpora (containing translations of the same source speeches) to identify patterns that could potentially be source language specific and eliminate them from the analysis, as it is assumed that universal features cannot be a result of interference. The only exception to that procedure will be made in the analysis of simplification, because the corresponding translation corpora are significantly larger than interpreting corpora which could influence the results of the tested parameters. It is believed that a feature may be an interpreting universal, if the values of the examined parameters are evenly distributed across all interpreting corpora proving that they are a homogenous group, where each

constituent resembles the others and remains at the same time significantly different from the corpus of native English speeches.

The micro analysis of interpreting style will be performed in a manner very similar to the macro analysis but will be significantly different in scope (see Figure 2 below). Its goal is to compare non-interpreted speech of two interpreters I1 and I2 with their interpretations from different source languages. The data used in micro analysis is much smaller, less balanced and limited to texts produced by two interpreters (see Section 2.2.2. and Section 2.2.3.). Simplification, explicitness and normalisation will be investigated in altogether four corpora of I1's texts: his original speech, interpretations from French, interpretations from Spanish and interpretations from German. The same features will be sought for in three I2's corpora: original speech, interpretations from French and interpretations from German. If it is found that all interpreter's texts are consistent with respect to one or more features and also deviate from the texts produced by the other interpreter then the feature will be considered a trait of that individual interpreter's style.

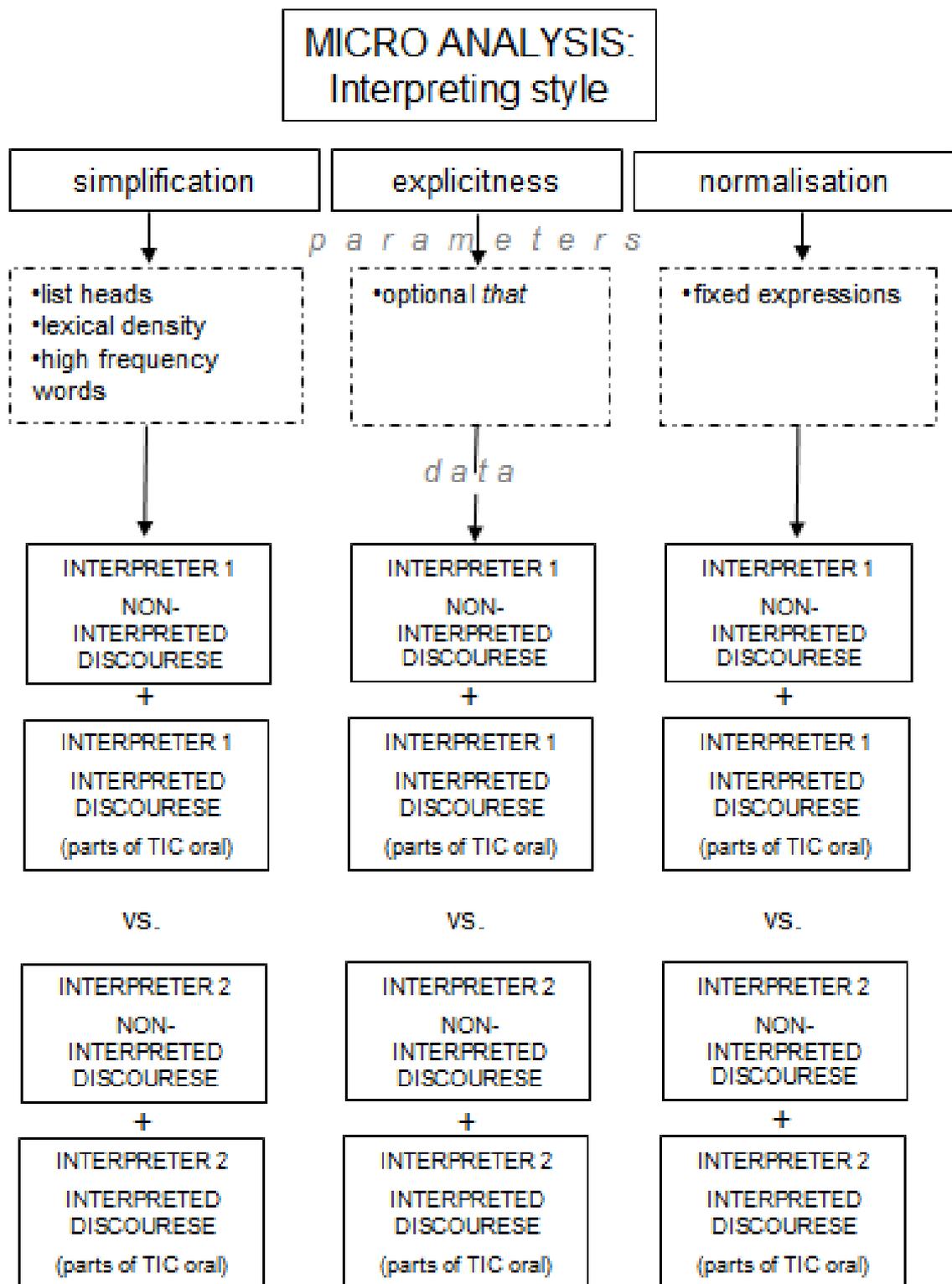


Figure 2. The structure of micro analysis of interpreting style depicting the investigated features, parameters and data sets.

2.2. Data

Investigations in the present study are carried out on two sets of data: Translation and Interpreting Corpus (TIC) and small corpora of texts produced by two interpreters. The former is used in the macro analysis of putative interpreting universals (Chapter Three) , while the other is examined in the micro analysis of interpreting style (Chapter Four).

2.2.1. Translation and Interpreting Corpus (TIC)

TIC is a monolingual comparable corpus comprising interpreted and translated texts as well as texts originally produced in English. It consists of two subsets: a subset of five oral subcorpora and a subset of five written subcorpora (see Figure 3). Probably the most important feature of TIC is that interpreted texts included in oral subcorpora and translated texts in written subcorpora originate from the same source texts i.e. speeches delivered in the European Parliament by either French, Spanish, German or Dutch MEPs (depending on the subcorpus). Since TIC is a monolingual comparable corpus, it has not been aligned and does not include the source speeches.

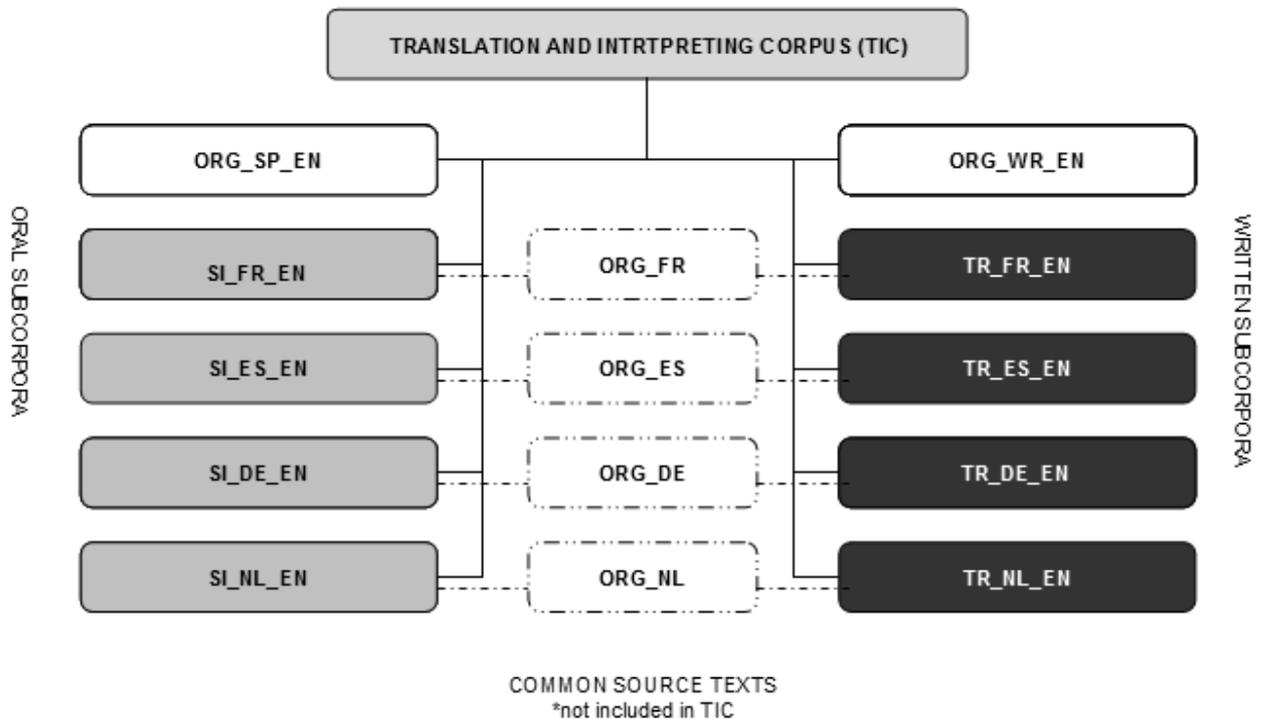


Figure 3. Translation and Interpreting Corpus.

The subset of oral subcorpora comprises transcriptions of speeches interpreted simultaneously into English from two Romance languages (French and Spanish) and two Germanic languages (German and Dutch) as well as a reference subcorpus of transcriptions of speeches given originally in English (semi-scripted or scripted but not read aloud).

The following corpora constitute the oral subset of TIC:

- (1) subcorpus of speeches interpreted simultaneously from French into English (SI_FR_EN)
- (2) subcorpus of speeches interpreted simultaneously from German into English (SI_DE_EN)
- (3) subcorpus of speeches interpreted simultaneously from Spanish into English (SI_ES_EN)

- (4) subcorpus of speeches interpreted simultaneously from Dutch into English (SI_NL_EN)
- (5) subcorpus of speeches originally spoken in English (ORG_SP_EN)

As demonstrated in Figure 4 the size of each of interpreting subcorpus is almost identical to facilitate their analyses and subsequent comparison. All oral subcorpora slightly exceed 50,000 words, which makes the total size of the oral subset of Translation and Interpreting Corpus (TIC) exceed over 250,000 words.

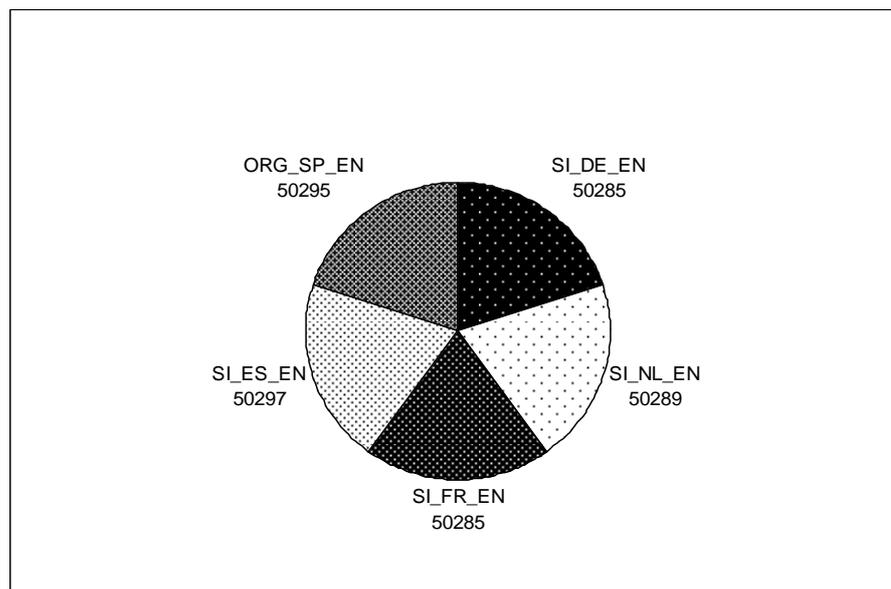


Figure 4. Composition of oral subcorpora of TIC.

The subset of written subcorpora consists of written translations of speeches into English from two Romance languages (French and Spanish) and two Germanic languages (German and Dutch) as well as a reference subcorpus of speeches read out in English by native English MEPs. The written subset of TIC is built from the following components:

- (6) subcorpus of speeches translated from French into English (TR_FR_EN)
- (7) subcorpus of speeches translated from German into English (TR_DE_EN)
- (8) subcorpus of speeches translated from Spanish into English (TR_ES_EN)
- (9) subcorpus of speeches translated from Dutch into English (TR_NL_EN)
- (10) subcorpus of speeches originally written in English (ORG_WR_EN)

As shown in Figure 5 translation subcorpora and the reference subcorpus of speeches written and read out by native English MEPs have different word counts. The most important selection criterion was to include in the respective translation subcorpora translations of all the speeches that constitute respective interpreting subcorpora to enable the subsequent comparison of interpreted and translated texts that share the same source texts. Translations turned out to be longer, hence subcorpora in the written subset of TIC vary in size.

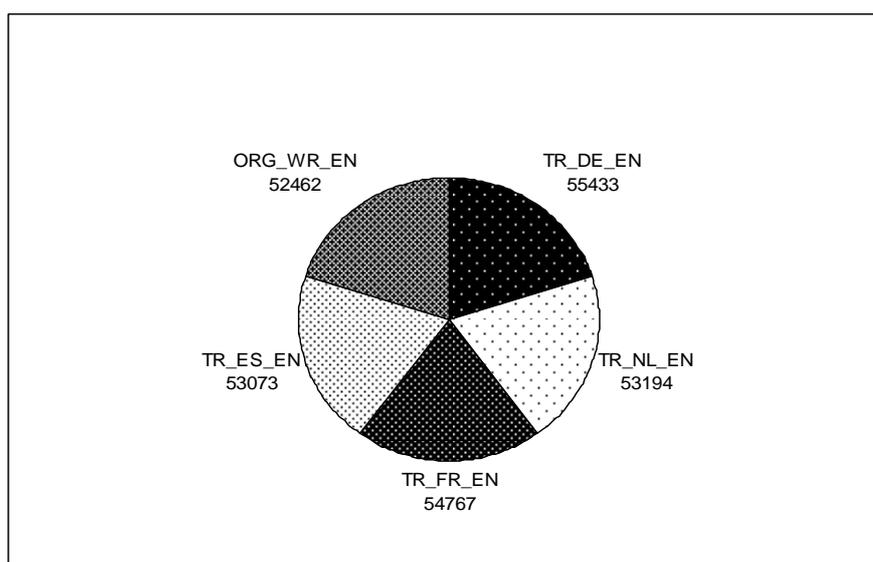


Figure 5. Composition of written subcorpora of TIC.

Since each corpus is bigger than any of the oral subcorpora the total size of the written subset of TIC exceeds 260,000 words.

2.2.1.1. TIC: Oral subcorpora

There are few publicly available resources offering access to authentic interpreting data. The Media Library of the European Parliament website, however, provides quality recordings of plenary sessions of the European Parliament together with their interpretations into all official languages of the European Union. The material is highly homogenous: speeches are given always in the same formal and institutionalised setting, speakers speak mostly in their mother tongue, interpretation is provided by the EP

interpreting staff or EU accredited professionals and in most cases also into the mother tongue. To increase the chances of meeting the latter criterion only interpretations from French, German, Spanish and Dutch into English have been included in TIC since these four languages are best represented in the English interpreting booth at the European Parliament.

At the outset of the compilation process TIC was supposed to become a corpus bearing the following characteristics:

- balanced representation of approximately 25 interpreters per language interpreting directly into their mother tongue (i.e. without relay)
- selected interpreters should have other TIC source languages in their combination (the more the better)
- relatively balanced coverage of different topics
- balanced proportion of interpretations of speeches that were originally spoken and read out

Achieving a perfect balance in a corpus consisting of authentic, non-experimental data is, however, very difficult and many of the initial compilation criteria had to be compromised.

The ideal design of oral subcorpora of TIC would require each of the interpreting subcorpus to consist of equal number of speeches on similar topics, equal in length, interpreted by an equal number of different interpreters. The number of interpreters should be big enough to ensure a variety of interpreting styles and small enough to allow for a comparison between individual interpreters (with the subcorpus size limited to approximately 50 000, a higher number of interpreters would shorten the total word count of text per interpreter and make such a comparison impossible). In reality it is very difficult to find a large number of speeches interpreted by one interpreter. Even in the Media Library of the European Parliament, which employs an extremely high number of interpreters it poses a problem, since the interpreters are summoned to different meetings and interpret at the recorded ones very irregularly. The overriding goal for the compilation of TIC was to enable the analysis of ‘global’, recurring patterns on a balanced collection of interpreted texts as well as the idiosyncratic variation in interpretations by individual interpreters from different source languages into one target language. Thus it was decided

that the most important criterion of all is to achieve in each interpreting subcorpus a compilation of texts interpreted by a relatively balanced number of individual interpreters, who would ideally have more than one of TIC source languages in their combination (the later prerequisite allows for a comparison of texts interpreted by one interpreter from different source languages). It was considered extremely important that performance of one active interpreter does not heavily dominate particular interpreting subcorpus. It turned out that to satisfy this requirement to a certain extent all the remaining criteria had to be compromised.

In the subcorpora of interpretations from French, German the main criterion of balance was easier to reach because there is a high number of interpreters in the English booth interpreting from those languages. Hence SI_FR_EN and SI_DE_EN consist of interpretations delivered by 28 and 24 interpreters respectively. It was impossible to achieve the target number of speeches in case of Spanish and especially in case of Dutch, which comprise the work of 20 and 16 interpreters. Interpretations from Spanish and Dutch are carried out usually by the same interpreters. The compilation of the Dutch corpus was most challenging due to a relatively low number of Dutch MEPs and a limited (when compared to other main languages) number of interpreters with Dutch in the English booth. As a result the latter interpreting subcorpora are certainly less varied. Also it was unavoidable that the total word count of texts delivered by individual interpreters be uneven: some interpret more frequently than others but in the end all subcorpora are relatively balanced in terms of the percentage of texts of individual interpreters in the sense that no subcorpus is completely dominated by one or two interpreters (see Figure 6).

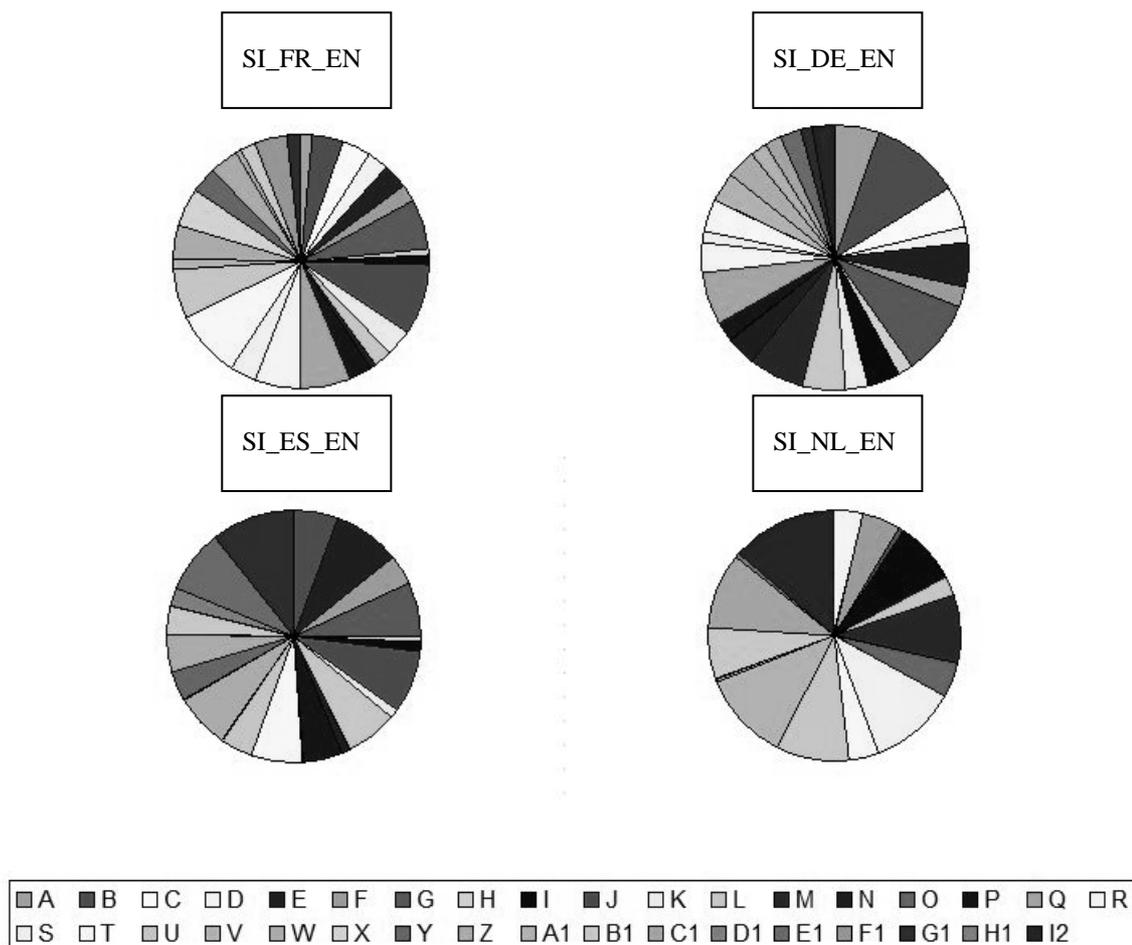


Figure 6. The distribution of interpretations carried out by individual interpreters across respective subcorpora

Another major goal of TIC, which distinguishes it from other available interpreting corpora was that it contain interpretations delivered by the same interpreter from different source languages (see Figure 7). This objective has been met in most cases but two: Dutch and Spanish. Were it not for interpreters O, D1 and H1 it would be impossible to reach the target 50000 word size of both interpreting subcorpora.

The length of the speeches in all TIC oral subcorpora ranges between approximately 100 words to 1000 words. The majority of recordings are on average one and a half to three minutes long and the longest do not exceed seven minutes. The word count depicted in Figure 7 shows the total number of words per interpreter and not the length of individual speeches.

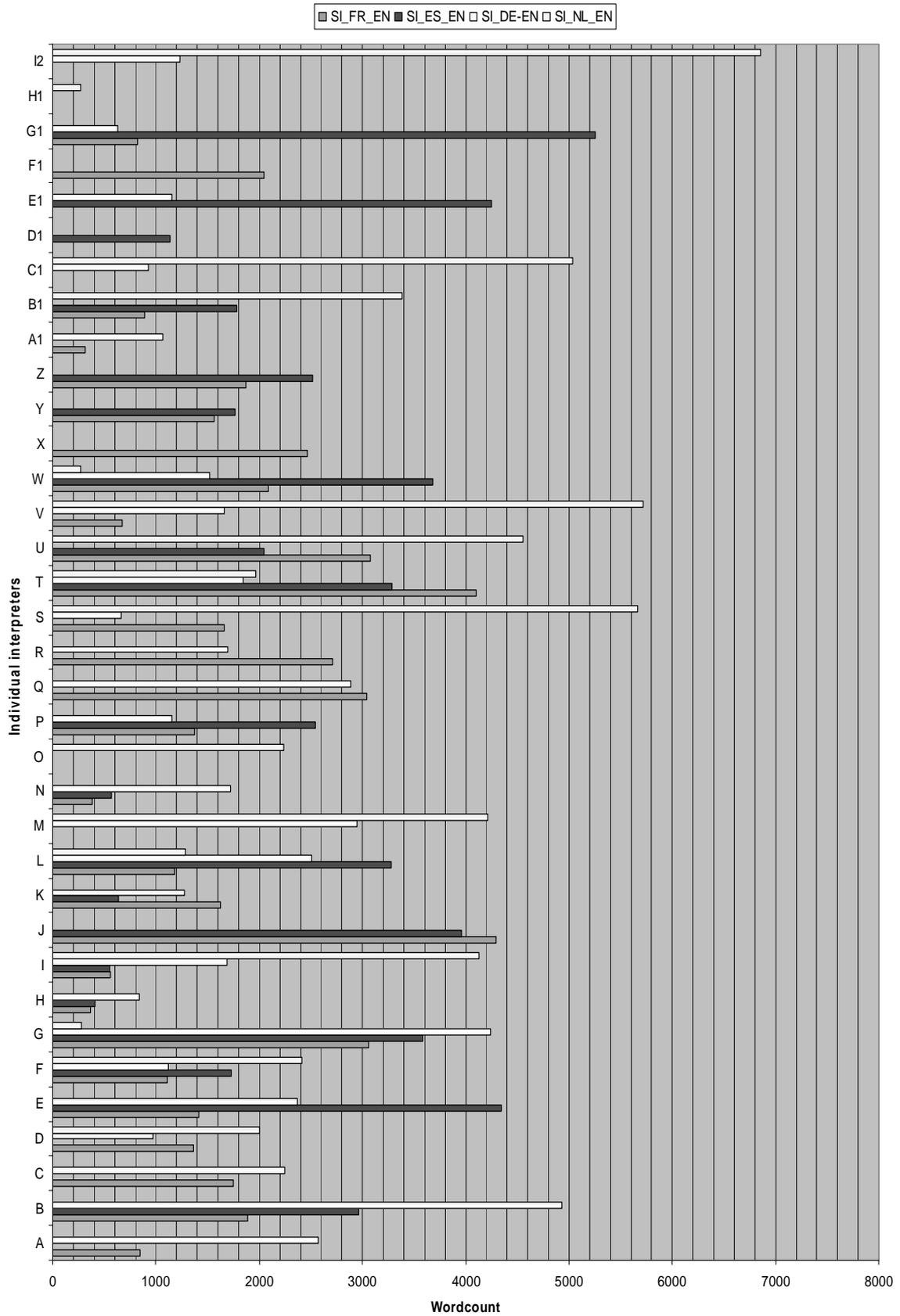


Figure 7. The total length of texts interpreted by individual interpreters from different languages expressed in number of words.

Since information on the author of interpretation is not available at the Media Library of the European Parliament website, in order to achieve this goal it was necessary to resort to a speaker recognition tool, which was used to distinguish between the voices of different interpreters to make sure that the corpus is balanced. The software used in the present study was RecoMadeEasy® Speaker Recognition: a language and text-independent speaker recognition system which has been developed to work in different environments. It is based on a speaker recognition technique that in very general terms uses the vocal characteristics of an individual's voice to identify and verify the person (<http://www.recotechnologies.com>)(date of access: 11 November 2011).

The speeches in all oral subcorpora cover a wide range of topics i.e. human rights, politics, industry, international affairs, economy and finance, law, environment, health and agriculture (see Figure 8). Although the Multimedia Library of the European Parliament provides recordings of debates on different topics, it was impossible to include equal number of speeches on a given topic in all respective subcorpora. The priority in the compilation of the corpus was given to a relatively balanced representation of different interpreters and to interpreting provided without relay (a pivot language). It is noticeable that certain topics are underrepresented in particular subcorpora which might have been caused by numerous reasons e.g. the subcorpus of speeches interpreted from Dutch into English contains only one speech on agriculture. This might have been caused by different factors, such as the Dutch MEPs speaking a foreign language during debates devoted to agriculture, no Dutch MEPs expressing their views on this topic or interpretation into English provided via relay (e.g. from French), which could not be included in the corpus.

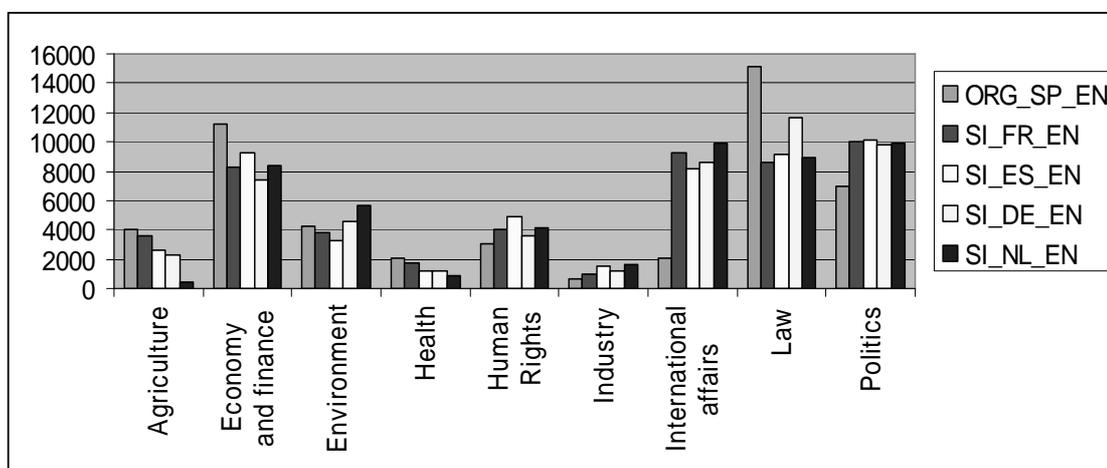


Figure 8. Topics in oral subcorpora of TIC

Finally, an important factor was the mode of delivery of the original speech. When interpreting a written speech the interpreter is not expected to produce an interpretation that would in all characteristics resemble a written text, as the nature of interpretation is by definition oral. But the mere exposure to the qualities of a written text may in certain aspects influence the interpreter's output. The degree and scope of such impact depends on many factors, among other: the character of the written text, interpreter's attitude and the context of interpretation. This factor should be considered in the analyses to follow, especially the ones focused on simplification.

Ideally the corpus should contain 50% of interpretations of speeches originally spoken (spontaneous or semi-scripted) and the same share of texts that constitute an interpretation of what was originally read out. This criterion was difficult to meet because the vast majority of MEPs read their speeches aloud. In the end, French and Dutch source speeches of the interpretations included in TIC were more frequently read out, German ones were mostly spoken and the subcorpus of interpretations from Spanish is balanced with this respect (see Figure 9).

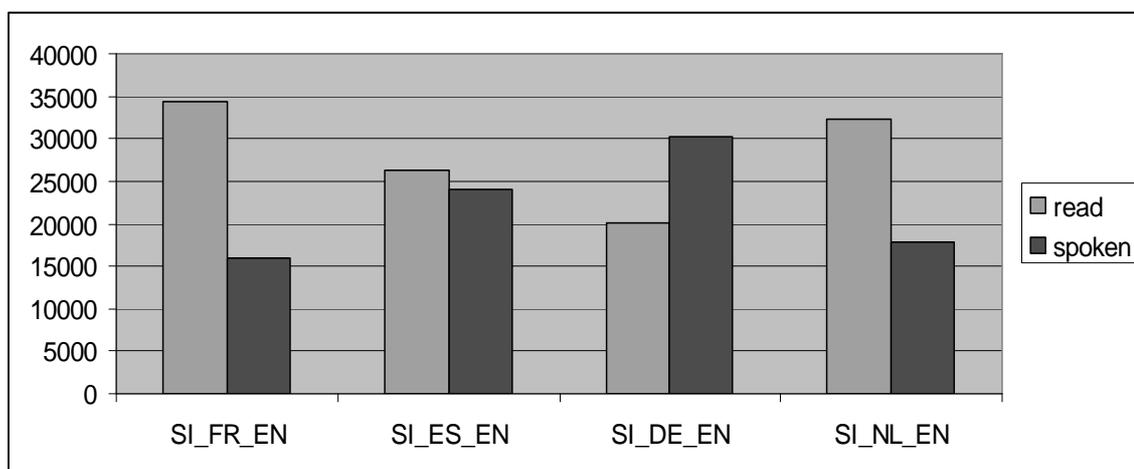


Figure 9 Mode of delivery of original speeches across respective interpreting subcorpora (expressed in number of tokens).

2.2.1.2. TIC: Written subcorpora

The subset of written subcorpora is a collection of speeches translated into English from two Romance languages (French and Spanish) and two Germanic languages (German and Dutch) as well as a reference subcorpus of speeches written originally in English and read out by English MEPs. All materials used in the corpus were obtained from the official website of the European Parliament. Original English speeches given by MEPs representing the United Kingdom and Ireland are available at the deputies' official EP websites. These materials comprise ORG_EN_WR subcorpus. Similarly, the same source provides English translations of speeches given in French, German, Spanish and Dutch by MEPs from these Member States, which, in turn, were used in compilation of TR_FR_EN, TR_DE_EN, TR_ES_EN, TR_NL_EN subcorpora.

The speeches in all written subcorpora cover a wide range of topics i.e. human rights, politics, industry, international affairs, economy and finance, law, environment, health and agriculture (see Figure 10). The topics in translation subcorpora mirror exactly the proportion of topics in interpreting subcorpora because both the interpretations and translations share the same source texts. The limitations in the scope and representation of the topics therefore also correspond to the ones concerning the oral subcorpora (see section 2.2.2).

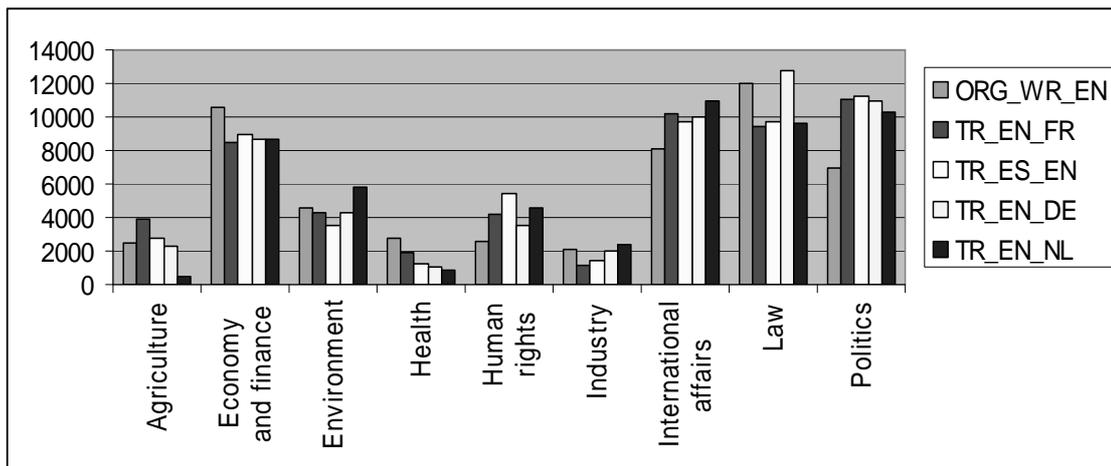


Figure 10. Topics in written subcorpora of TIC

Defining a reference corpus for translation subcorpora poses a methodological problem in the very specific case of TIC, since the texts in translation subcorpora constitute written translations of transcripts of speeches that have originally been spoken or read out. The texts in translation subcorpora are a unique mix of both oral and written discourse. It is thus very difficult to determine the type of native texts (produced originally in English) that could share exactly the same characteristics.

Transcripts of speeches spoken (not read out) by the native English MEPs are by definition oral texts using spoken language rather than written because the transcription process does not enforce any linguistic features of written texts on them.

For the lack of a better alternative TIC translation subcorpora shall be compared with written speeches read out in the European Parliament by native English MEPs. In theory, such texts would ideally combine features of both written and oral language since their authors should respect the fact that the texts are delivered orally. In practice, however, different authors of the so called ‘written to be read’ speeches show a very liberal attitude towards these conventions and the final texts may not differ significantly from written discourse. As depicted in Figure 11 the subcorpus of speeches written and read out by the native English MEPs is relatively distant on the literacy – orality axis from e.g. the subcorpus of translations of German speeches, the greatest proportion of which were originally delivered orally in unscripted or semi-scripted manner.

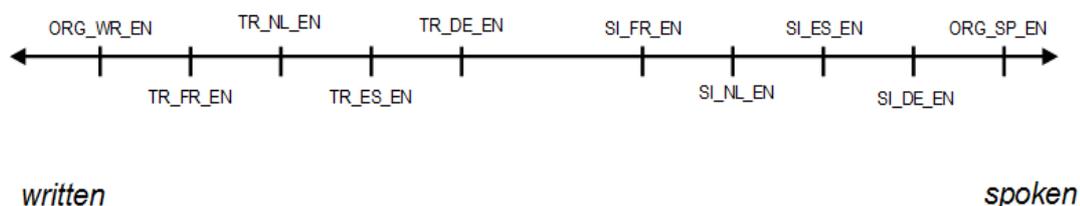


Figure 11. Location of TIC subcorpora on the literacy – orality axis.

Consequently, in the following analysis one must bear in mind that certain discrepancies between respective written subcorpora of TIC might be ascribed to not only the differences between native English speeches and translations, but also to the differences

between written and oral modalities. Features of both the former and the latter, although present in all written subcorpora of TIC, may not be evenly distributed.

Such a compromise has only been allowed in this study, since its overriding aim is to verify if the hypothesised translation universals are in any way reflected in interpreting and not to test them once again on translations. Subcorpora of translations are used in this study predominantly as a control set, to help determine whether specific results of analyses at different stages may or may not be ascribed to language transfer.

2.2.1.3. Data storage and transcription

First, all audio recordings of the speeches were recorded from the Multimedia Library of the European Parliament. Since it is possible to download only the video with the voice of the original speaker, each interpretation was recorded in a separate audio file. Spoken data of TIC were transcribed according to the guidelines similar to the ones applied in EPIC (European Parliament Interpreting Corpus).

EPIC transcription conventions are selective. They are limited to some aspects of different transcription levels: linguistic, paralinguistic and extra-linguistic. The idea of EPIC researchers “was to produce a basic transcript, to which further levels of annotation could be easily added if and when needed” (SSLMIT 2004), which is generally in line with the present project.

According to EPIC translation conventions linguistic information i.e. orthography, punctuation, segmentation, spelling should be transcribed with respect to the following rules:

All the words spoken by both speakers and simultaneous interpreters are transcribed (orthographic transcription). There are no punctuation signs in the transcripts, as they could be misleading and create problems in automatic analysis. Transcribed texts are segmented in units of meaning, on the basis of the speaker's intonation and syntactic information in the sentence involved. The double bar sign // is used to indicate the end of each segment [...].Spelling conventions follow the standards applied in EU official documents. These indications can be found in the Interinstitutional Style Guide which is available on the European Parliament website for all the official languages of the Union. Figures, dates and percentages are fully spelt out (SSLMIT 2004).

Transcription conventions applied in EPIC take into account only a small number of paralinguistic features: truncated words, mispronunciations and pauses. Their transcription is shown in Table 1. Transcription conventions in TIC adapted from EPIC .

Table 1. Transcription conventions in TIC adapted from EPIC.

Speech feature	Example	EPIC transcription conventions	TIC transcription conventions
word truncations	propo pro posal	propo- proposal /pro_posal/	proposal <propo-> proposal <pro_posal>
pronunciation disfluencies	parloment	Parliament /Parloment/	parliament <parloment>
filled pauses		ehm	<ehm>
empty pauses		...	<...>
numbers	532	five hundred and thirty- two	five hundred and thirty- two
figures	4%	four per cent	four per cent
dates	1997	nineteen ninety-seven	nineteen ninety-seven
unintelligible		#	#
units		based on syntax and intonation	based on meaning and intonation

Transcription conventions used in TIC differ slightly from what has been presented above. All truncated words have been transcribed using the correct spelling and the way they have actually been pronounced has been placed in triangular brackets. Similar procedure has been applied to strings of words considered false starts and filled or empty pauses. Thanks to the use of triangular brackets, such phenomena were not counted as separate types or additional tokens and completely disregarded in corpus analysis.

Since EPIC was also automatically POS-tagged, mispronounced words and words with “an internal truncation were first ‘normalised’ and then transcribed as they were actually spoken. For truncations we use the - symbol at the end of the truncated word” (SSLMIT 2004). The same procedure was applied in TIC, since it might be POS-tagged in the future.

EPIC transcription conventions take also pauses into account. They are not measured objectively, but transcribed accordingly to the transcribers’ perception. Silent pauses are marked with “(..)” and filled pauses with “(ehm)” but no details are provided about their duration.

Extra-linguistic information is encoded in a form of headers, which provide the following details: speaker’s/ interpreter’s: gender, speaker’s country, source language, speech duration, length (in number of words), topic.

2.2.2. Individual interpreting corpora (parts of TIC)

Interpreting corpora of Interpreter 1 and Interpreter 2 are compiled of transcriptions taken from interpreting subcorpora of TIC.

There are three corpora comprising texts interpreted by Interpreter 1 from three different languages: French, Spanish and German into English. The corpora have the following sizes:

- I1_SI_FR_EN 1892 tokens (interpretations from French into English)
- I1_SI_ES_EN 2965 tokens (interpretations from Spanish into English)
- I1_SI_DE_EN 4929 tokens (interpretations from German into English)

As depicted in Figure 12 the three corpora vary with respect to the mode of delivery of the original speeches.

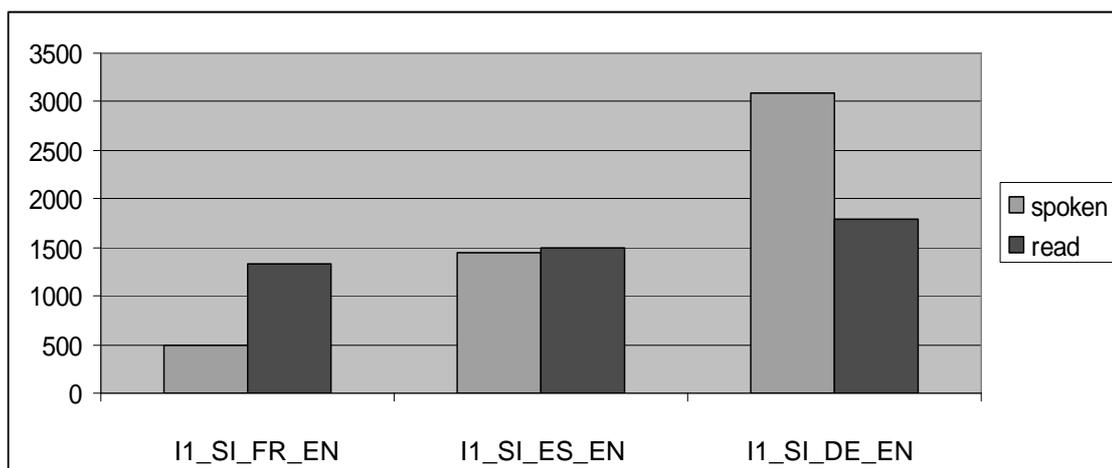


Figure 12. Mode of delivery of original speeches across respective corpora of Interpreter 1's interpretation (expressed in number of tokens).

There are two corpora comprising texts interpreted by Interpreter 2 from two different languages: French and German into English. The corpora have the following sizes:

- I2_SI_FR_EN 2044 tokens (interpretations from French into English)
- I2_SI_DE_EN 4428 tokens (interpretations from German into English)

As depicted in Figure 13 the two corpora vary with respect to the mode of delivery of the original speeches.

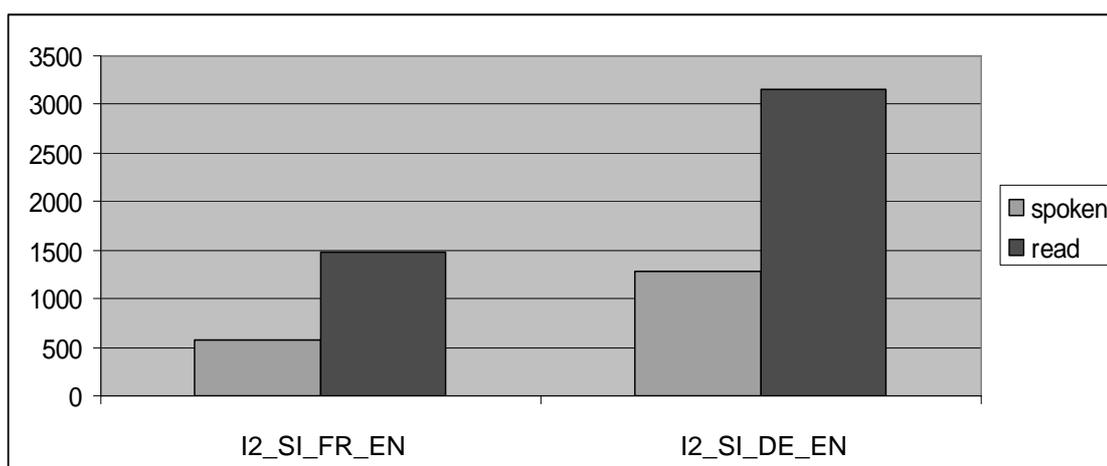


Figure 13. Mode of delivery of original speeches across respective corpora of Interpreter 2's interpretation (expressed in number of tokens).

The full corpora of Interpreter 1 and Interpreter 2 will be used in the investigation of their tendency to explicitness (see Section 4.3.) and normalisation (see Section 4.4.).

The tendency to simplification will be investigated on a specially prepared smaller corpora of text samples due to a great susceptibility of all three investigated parameters of simplification to corpus size. Each interpreting corpus tested in Section 4.2.1. consists of samples of four interpretations delivered by Interpreter 1 and Interpreter 2 of speeches that were read out at the European Parliament by four different source speakers. Samples include 60-word fragments from the beginning, ending and middle sections of the text. Thus each sample is 180 words long and each corpus counts 720 words. To increase the comparability of the corpora, samples were selected only from interpretations of speeches that were originally read out. Unfortunately an analogous comparison of interpretations of unscripted and semi-scripted speeches is not possible due to the lack of data.

2.2.3. Reference corpora of interpreters' spoken discourse

In the micro analysis the tendency of Interpreter 1 (I1) and Interpreter 2 (I2) to simplification, explicitness and normalisation in interpreting will be examined on a subcorpus of their speaking output to see if the linguistic patterns are in any way reflected in their non-interpreted speech.

The recordings took place in an informal setting and involved the interpreters answering the questions about the European Union.

Both interpreters are native speakers of English who work for the European Parliament on a regular basis and have both Romance and Germanic languages in their combination. They were recorded for approximately an hour. At the beginning they were first asked warm-up questions about their education and their interpreting career, which were followed by a set of questions on the European Union. The answers to the warm-up questions are not included in the corpora. The idea behind the questions regarding the European Union was to keep the interpreters talking on the topics related to the ones that they usually face in interpreting. The questions were treated as prompts and the interpreters were allowed to depart from the main topic, as long as the digressions were related to the European Union. The recordings were transcribed and compiled into two corpora I1_ORG_SP_EN of 5455 tokens and I2_ORG_SP_EN containing 5788 tokens. Those full compilations were used in Chapter Four in the analysis of explicitness (see Section 4.3.) and normalisation (see Section 4.4.).

Similarly to the corpora of Interpreters' I1 and I2 interpretations, the corpora of non-interpreted discourse were specially prepared for the analysis of simplification. Samples include 60-word fragments from the beginning, ending and middle sections of the monologues produced in answers to the questions. Thus each sample of a monologue contains 180 words long and each corpus counts 720 words.

2.3. Method for macro analysis: Interpreting universals

The aim of the macro analysis is to verify if simplification, explicitness and normalisation may be considered as interpreting universals. For that purpose a compilation of parameters based on previous research carried out in Translation Studies will be used to investigate prominent tendencies across interpreting and translation subcorpora of TIC.

2.3.1. Simplification

Simplification shall be investigated according to Laviosa's (1996) methodology. As described in section 1.4.1. Laviosa believed that compared to texts originally written in the target language translated texts demonstrate lower lexical density, higher proportion of high frequency words versus lower frequency words, high repetition of the most frequent words and that the list head of translated texts contains fewer lemmas (Laviosa 1998: 563). At the present stage of development of TIC it is possible to analyse only three of those aspects: lexical density, the proportion of high frequency words vs. low frequency words, and the percentage of the corpus represented in the list head. All those are measured in the subcorpora of original English speeches (spoken) versus speeches interpreted into English.

The first measure applied in the present study focuses on list heads i.e. the first hundred words on the frequency list of a corpus. It is examined how extensive a part of the corpus is represented in the list head. A higher value is symptomatic of a more repetitive language and therefore indicates lexical simplification.

Second method involves the analysis of lexical density expressed by the proportion of lexical words to function words in the corpus certifying to the level of informativeness of the investigated text. The number of lexical words is obtained by subtracting the number of function words from the total number of running words. Since TIC has not been POS-tagged the only possible way to calculate the number of function words in the corpus is to feed a list of grammatical words into the concordancing tool (Wordsmith 5.0). To assure consistency with Laviosa's (1996: 119) methodology the number of function words in the present study is always generated based on the same list of grammatical words identified by Stubbs (1996: 36-37) and used by Laviosa (1996: 119) in her investigation of translational English (see Appendix 1). It is believed that the more function words are found in a corpus, the lower lexical density and the more simplified the language.

Finally, the last analysis is based on the proportion of high frequency words to low frequency words calculated with reference to a list of 200 most frequent words in English. The list was produced for Laviosa by Stubbs (as referred to by Laviosa in 1996: Appendix 34) based on Collins Cobulid Bank of English. The analysis involves examining how large a percentage of the corpus the 200 words constitute (the list can be found in Appendix 2). It is assumed that the higher the percentage, the more simplified the text.

All of the above described measures indicate how lexically complex the examined texts are. It is assumed that texts characterised by the lowest lexical density and the highest proportion of most frequent English words are the most simplified. The same can be assumed if the list head represents a large part of a corpus.

Simplification is the only parameter in the macro analysis that will be investigated only on oral subset of TIC, i.e. only on the interpreting subcorpora and the reference subcorpus of native English speeches. The reason for such a decision is the fact that all of the above described parameters of simplification are exceedingly susceptible of corpus size and translation subcorpora are much larger than interpretation subcorpora and extremely uneven. It was believed that if the corpora were trimmed to equal sizes they would no longer constitute the exact reflection of the interpreting corpora and reported findings could only lead to greater confusion.

2.3.2. Explicitness

Explicitness hypothesis will be verified with a combination of methodologies developed by different scholars that have delved into explicitness and explicitation patterns in translation. Investigated parameters include the use of: optional *that* connective, linking adverbials and apposition markers.

The first analysis focuses on the use of optional *that* after reporting verbs. The methodology was initially applied by Burnett (1999) as well as Baker and Olohan (2000). The former analysed the occurrence of optional *that* after the reporting verbs *admit, believe, claim, hope, know, suggest*, whereas the latter focused on the patterns of *say* and *tell* which may also be optionally followed by *that*. The present study aims to investigate the use of optional *that* in interpreted, translated and native English texts with reference to the following reporting verbs i.e. *admit, believe, hope, know, suggest, say* and *tell*. Any instance where the verbs used in their reporting function are followed by *that* will be classified as a form of syntactic explicitness. Consistently with Olohan and Baker's (2000: 152) methodology, all TIC subcorpora will be searched for all the forms of the reporting verbs in question e.g. *tell, telling, tells, told*. All concordance lines irrelevant to the present analysis of the patterns of optional *that* will be discarded i.e. all instances including:

Intransitive uses, including fixed or semi-fixed expressions

“Please don’t tell”

Infinitival complements

You’ve every right to tell me to mind my own business.

Non-clausal complements

“and I don’t tell lies”

Tell him as little as possible.

Interrogative clauses

“Let’s go tell the others what we plan to do.”

I can’t tell you how sad it was.

Phrasal and prepositional verbs

There was not a noise but seemed to tell of danger.

It’s often tough to tell them apart.

Formulaic, lexicalised expressions; set phrases and idiomatic expressions

“Tell me, Dr Gilly, what did you see?”

“Well, I tell you, Alfred, it’ll be a relief to get it”

And let me tell you, coming off coke is the hardest damn ...

Truth to tell, it is not difficult to see why.

(Olohan and Baker 2000:153)

It is assumed that the highest prevalence of optional *that* in individual subcorpora indicates a high level of syntactic explicitness.

Analysis of optional *that* will be followed by an examination of linking adverbials. The list of adverbials subject to investigation has been selected by Kruger and van Rooy (2010: 10) in their pilot study of the features of non-literary translated language. Based on Biber et al. (1999: 875-879) they decided to examine the following linking adverbials: *as a consequence, as a result, consequently, hence, in consequence, therefore, thus*. The same set will be applied in the present study. It is anticipated that interpreted and translated texts will rely more heavily on the use of linking adverbials.

Finally, the last investigated parameter of explicitness is the occurrence of apposition markers. The present analysis is based on the methodology adopted by Mutesayire (2004), who focused on reformulation phrases with explicitation function i.e. apposition markers, such as: *that is, that is to say, to be (more) precise, to be (more)*

specific, to be exact, namely, in other words. Since *to be (more) precise, to be (more) specific, to be exact* do not occur in the spoken subcorpora of TIC, they will not be analysed here. While most of the phrases are unlikely to perform other functions than assumed in this study, occurrences of *that is* will have to be meticulously examined manually. The phrase is relatively rarely used as an apposition marker and its more standard functions are bound to outnumber such occurrences and the majority of concordance lines are likely to be discarded.

It is predicted that apposition markers should be better represented in subcorpora of interpreted and translated texts as opposed to text originally produced in English.

2.3.3. Normalisation

In line with Baker's (1996: 176-7) definition of normalisation perceived as the tendency to opt for and even exaggerate linguistic patterns typical of the target language, this study attempts to evaluate normalisation in interpreted and translated texts using two different methods. The first involves a comparison of the frequency of repeated strings of words, i.e. lexical bundles common for all compared corpora, while the second focuses on fixed expressions most typical of a formal register of spoken English.

Lexical bundles are strings of words that recur in the analysed corpus and do not necessarily bear any semantic meaning. In other words, what makes a string of words a lexical bundle is its repetition and not the semantic meaning. Lexical bundles are also referred to as *n*-grams, with the letter *n* denoting the number of elements in the bundle. In the present study bundles consisting of three elements are investigated and are referred to as trigrams. Terms *lexical bundles* and *trigrams* are used in this study interchangeably.

The lexical bundles for the analysis of normalisation have been selected very carefully from two reference corpora: ORG_SP_EN (unscripted or semi-scripted speeches produced orally by native English MEPs) and ORG_WR_EN (speeches read out by native English MEPs) and compiled into two separate lists of trigrams to be tested on the interpreted and translated subcorpora. After the list of all trigrams in a given reference subcorpus was generated it was limited only to the ones that did not contain subject-specific or proper nouns and occurred at least 10 or 7 times in the reference subcorpora (10 in ORG_SP_EN and 7 in ORG_SP_WR). It was presumed that all less frequent trigrams

would not occur at all in interpreting/ translation subcorpora. Finally, the list was limited to 20 trigrams that occurred in all oral/written subcorpora of TIC 10 or 7 times (in oral and written respectively) and were most frequent in the reference subcorpus. The following lists of trigrams were generated based on ORG_SP_EN and ORG_WR_EN:

Table 2. Trigrams selected for investigation.

ORG_SP_EN	ORG_WR_EN
<i>and I think</i>	<i>and it is</i>
<i>at the moment</i>	<i>be able to</i>
<i>be able to</i>	<i>I do not</i>
<i>first of all</i>	<i>I would like</i>
<i>going to be</i>	<i>in order to</i>
<i>I think it</i>	<i>is not a</i>
<i>I think that</i>	<i>it is a</i>
<i>I think we</i>	<i>on behalf of</i>
<i>I would like</i>	<i>one of the</i>
<i>in terms of</i>	<i>part of the</i>
<i>is going to</i>	<i>that we can</i>
<i>make sure that</i>	<i>that we have</i>
<i>that we have</i>	<i>the fact that</i>
<i>there is a</i>	<i>there is no</i>
<i>this is a</i>	<i>this is a</i>
<i>to have a</i>	<i>to ensure that</i>
<i>to make sure</i>	<i>we do not</i>
<i>we have to</i>	<i>we have to</i>
<i>we need to</i>	<i>we need to</i>
<i>would like to</i>	<i>would like to</i>

The frequency of lexical bundles presented in Table 2 will be investigated in oral and written subcorpora of TIC. It is anticipated that despite the fact that they were generated based on the reference subcorpora their frequency in interpreted and translated texts will be significantly higher, which would point to a tendency to normalisation.

The second methodology for investigating normalisation focuses on the frequency of the most common fixed expressions. It is necessary to first establish which register is the most appropriate platform of comparison for TIC subcorpora and which fixed expressions are most typical of its spoken variety. Since in the European Parliament it is the British and the Irish MEPs who use English as their native language, such a list would ideally be prepared based on a large corpus of speeches delivered in the British and the Irish Parliament. Alternatively, a list of such expressions could be generated based on a large subcorpus of political speeches in a spoken subcorpus of a national corpus of English, preferably contemporary British English, as it most probably is the dominant variety spoken in the European Parliament (among the native speakers of English). Since, developing a list of most frequent fixed expressions (not merely n-grams) based on any of the above would be extremely laborious and could constitute an object of research on its own, this study utilises a list developed for a different research project and customised to fit the purposes of the present analysis.

The selection of an adequate reference list of most frequent fixed expressions posed a huge methodological problem. Literature on the subject focuses rather on providing a description, classification and examples of use of different types of idioms, collocations and alike, but rarely includes the actual frequency list of fixed expressions used across different registers. Finding a list of most frequent fixed expressions characteristic of spoken formal register and not merely conversational English was therefore very difficult. In the end, the expressions chosen for the analysis of normalisation in the oral and written subcorpora of TIC have been elicited from the list of most frequently used fixed expressions in the Corpus of Spoken Professional American English (Liu 2003: Appendix C). Although such a choice may appear controversial, it has been selected as the best available choice for a number of reasons. Firstly, the Corpus of Spoken Professional American English (CSPAEE) is a compilation of transcripts of meetings at academic institutions and professional organisations and White House press briefings i.e. discourse used in fairly formal settings when compared to other relatively new spoken subcorpora of English e.g. the spoken subcorpus of the Corpus of Contemporary American English (COCA) that consists of unscripted conversation on TV and radio shows. Secondly, CSPAEE consists of recordings collected between 1994 and 1998 i.e. more recently than the spoken subcorpus of the British National Corpus (BNC) which was compiled between 1980s and 1993 and no available list of the most frequently used fixed expressions has been

compiled based on it, nor based on any other corpus of English political speeches. Thus the list based on CSPAE seems to be more valuable for the present study than other available lists developed for general spoken English corpora (e.g. Shin – Nation's (2008: 346-348) list of most frequent collocations in spoken English based on the BNC spoken subcorpus, which includes very colloquial items). Finally, since the expressions used in this study are limited only to the most frequent ones, it is believed that, although their rank on frequency lists generated for different varieties of English may differ, the set as such, would probably remain the same.

Liu (2003: Appendix C) reports that in his investigation he included three different types of idioms identified and classified after Fernando (1996: 32) as pure (e.g. *kick the bucket*), semi-literal (e.g. *fat chance*) and literal (e.g. *in sum*). When in doubt about whether an expression could be considered an idiom, the author consulted four major contemporary English idiom dictionaries and three dictionaries of phrasal verbs³. Such is the character of the list of approximately 300 expressions found to be the most frequent in CSPAE. However, as pointed out by Moon (1998: 2) fixed expressions and idioms are two very problematic terms. In Anglo-American traditions all fixed expressions are usually treated as idioms and such seems to be the case with Fernando's definition applied by Liu. The present study is guided by the definitions adopted by Moon (1998: 2), who classifies idioms as a subset of a broad category of fixed expressions i.e. holistic units of two or more words including alongside idioms also frozen collocations, grammatically ill-formed collocations, proverbs, routine formulae, sayings and similes. It is believed that phrases selected from Liu's study for the present investigation are in fact fixed expressions other than pure idioms that are here understood as semi-transparent and opaque expressions such as *spill the beans*.

The methodology of the current research requires the development of a stop list that could be used as a filter for a list of all 2-5 word clusters recurring in the examined oral and written subcorpora of TIC. This would allow for automatic extraction of fixed expressions from each of the respective subcorpus. To facilitate the process, the list obtained by Liu has been significantly shortened. First, phrases containing verbs and pronouns were eliminated

³ Cambridge International Dictionary of Idioms (1998), Cambridge International Dictionary of Phrasal Verbs (1997), Longman American Idioms Dictionary (1999), NTC's American Idioms Dictionary (Spears, 1994), NTC's Dictionary of Phrasal Verbs and Other Idiomatic Verbal Phrases (Spears, 1993), and Oxford Idioms Dictionary for Learners of English (2001) and Oxford Phrasal Verbs Dictionary for Learners of English (2001).

to narrow down the set only to those expressions, which are never inflected. A preliminary set was limited to 100 most frequent idioms. The final list was determined by the frequency of occurrences of the investigated expressions in the subcorpus of speeches originally spoken in English (ORG_SP_EN), which is in this study believed to reflect the spoken register of English used at the European Parliament by the native speakers. The additional condition was that all the examined phrases occur in all interpreting subcorpora at least once. Hence, the expressions with less than 7 occurrences in the corpus of native English speeches (ORG_SP_EN) and not included in the interpreting subcorpora have been eliminated. Finally, to make sure that none of the phrases are typical of the American variety only, all were examined in the Longman Dictionary of Contemporary English (date of access: 25 November 2011), which provides information on vocabulary items that are specific of other than the British variety. None of the selected expressions was marked as typical of American English or any other variety. Thus the final stop list assumed the following shape of sixteen items:

as far as, as to, as well, as well as, at all, first of all, in fact, in favour of, in order to in terms of, kind of, of course, on behalf of, once again, sort of, with regard to

It is understood in the present study, that overrepresentation of the above expressions in a subcorpus of interpreted discourse shall signify a tendency to normalisation. Since the translated and interpreted texts in oral and written subcorpora of TIC originate from the same set of source texts delivered orally (i.e. the same set of speeches originally given in French, Spanish, German and Dutch) it is moreover expected that any tendency revealed in the oral subcorpora of TIC with respect to the set of fixed expressions assumed to be most typical of the spoken register should be reflected also in the written ones.

2.4. Method for micro analysis: Interpreting style

The investigation of interpreting style needs to be adjusted to the size and make-up of the corpora used in micro analysis, as not all parameters can be delved into at such small scale. All parameters of simplification will be tested on a special selection of samples extracted

from individual interpreting corpora (see Section 2.2.2.) and reference corpora of interpreters' spoken discourse (see Section 2.2.3.). The analysis of explicitness and normalisation will be carried out on the full compilation of the two sets of corpora, but not all parameters will be examined due to low incidence of lexical items crucial for such a comparison. Hence explicitness will be analysed with the frequency of optional complementizer *that* and normalisation only with most frequent fixed expressions.

2.4.1. Simplification

The investigation of simplification in interpreting style will employ the same parameters that are utilised in the macro analysis of interpreting universals (see Section 2.3.1.) i.e. the calculation of lexical density, the percentage of the corpus covered by list head and the percentage of the most frequent words.

Given the uneven size of respective small corpora all the results necessitate normalisation since the parameters used are particularly susceptible to corpus size.

Calculation of the percentage of the corpus covered by list head is particularly problematic. The problem lies in the fact that the bigger the corpus, the greater the repetition of the list head. Since due to a very small size of the corpora compared in the course of micro analysis of interpreting style the potential differences in number of tokens could also affect the incidence of function words and high frequency words. Therefore only for the calculation of the three parameters of simplification all respective corpora had to be adjusted to the even size of 720 words. Such a solution results in a significant loss of data, but allows to generate comparable results. It has to be borne in mind that a 720 word-long text is not a representative sample and therefore all results should be treated as indicative only.

The three parameters of simplification will be tested altogether on seven corpora; four corpora of texts produced by Interpreter 1:

- I1_ORG_SP_EN 720 tokens (non-interpreted discourse)
- I1_SI_FR_EN 720 tokens (interpretations from French into English)
- I1_SI_ES_EN 720 tokens (interpretations from Spanish into English)
- I1_SI_DE_EN 720 tokens (interpretations from German into English)

and three corpora of texts produced by Interpreter 2:

- I2_ORG_SP_EN 720 tokens (non-interpreted discourse)
- I2_SI_FR_EN 720 tokens (interpretations from French into English)
- I2_SI_DE_EN 720 tokens (interpretations from German into English).

The results of the analysis of each parameter of simplification will be presented in a separate subsection, compared and discussed in Section 4.2.1. Also each interpreting corpus will be compared in a χ^2 test with the respective corpus of non-interpreted discourse.

2.4.2. Explicitness

Only one of the parameters of explicitness employed in the analysis of interpreting universals will be used to verify if individual interpreters differ with respect to this tendency i.e. the frequency of optional complementizer *that* after reporting verbs (in line with the requirements described in Section 2.3.2.). The parameter will be tested on all seven full-sized corpora of texts produced by Interpreter 1:

- I1_ORG_SP_EN: 5455 tokens (non-interpreted discourse)
- I1_SI_FR_EN: 1892 tokens (interpretations from French into English)
- I1_SI_ES_EN: 2965 tokens (interpretations from Spanish into English)
- I1_SI_DE_EN: 4929 tokens (interpretations from German into English)

and Interpreter 2:

- I2_ORG_SP_EN: 5788 tokens (non-interpreted discourse)
- I2_SI_FR_EN: 2044 tokens (interpretations from French into English)
- I2_SI_DE_EN: 4428 tokens (interpretations from German into English).

The results of the analysis will be discussed in Section 4.3. They will not be subject to any statistical test due to a limited number of occurrences and will only be treated as indications of possible tendencies. The remaining two parameters involving the analysis of the incidence of linking adverbials and apposition markers will not be tested due to insufficient size of the tested corpora.

2.4.3. Normalisation

Similarly to explicitness, the full analysis of the tendency to normalisation among two individual interpreters based on two parameters used in the examination of TIC is impossible due to the size of respective corpora and therefore must be limited to only one parameter, namely the investigation of most frequent fixed expressions characteristic of spoken professional English. Like the analysis of explicitness, this particular parameter will be tested on the full size of all corpora (see the list in Section 2.3.3. above). The observed frequencies will be normalised (per 1000 words), compared and discussed. Due to exceptionally low frequencies the trends will not be verified by any statistical test and will only be treated as indicative results.

2.5. Measure of statistical significance

All of the above methods of analysis are carried out to point to similarities and dissimilarities between the examined subcorpora. Obtained results may, in turn, reveal certain trends and patterns and to evaluate whether or not these are significant one has to reach for a test of statistical significance. For such estimation Lewandowska-Tomaszczyk (2005: 118) and Oaks (1998: 24) suggest chi-square test as the most appropriate non-parametric test allowing to compare in a display table frequencies discovered experimentally with those expected based on a theoretical model and in this way establish if the frequencies in the two populations differ significantly. The null hypothesis is a starting point to calculate the value of χ^2 (Chi squared):

If we employ the null hypothesis that there is no difference between the frequencies found in each category, the first step is to decide what the frequencies would have been if there were no relationship between category and frequency. In such a case, all the frequencies would be the same, and equal to the sum of the frequencies in each cell, divided by the number of categories (Oaks 1998: 24).

The value of χ^2 is therefore calculated according to the following formula (where O stands for observed frequencies, E for expected frequencies):

$$x^2 = \sum \frac{(O - E)^2}{E}$$

The significance of the chi-square value is assessed based on the number of degrees of freedom expressed as a product of the number of columns in the contingency table minus one and the number of rows minus one⁴. Chi-square value, on the other hand, allows to establish the p-value that points to the degree of probability that the generalisations inferred from the results yielded from the sample are wrong. For the purpose of this study it is assumed that the threshold of tolerance of an error be lower than 0.05 (p<0.05), which is consistent with the standards applied in many fields of research (Oakes 1998: 9).

In the following analyses interpreting and translation corpora shall be compared to reference subcorpora of speeches originally produced in English. As a rule a chi-square test will be run to verify the statistical significance of the difference between the values characteristic of the original speeches and the interpreting/ translation corpora and the χ^2 and p-values will be presented in tables just below raw frequencies or percentage values. All χ^2 tests will be computed on raw counts only. In the analyses of two parameters in micro analysis of interpreting style (see Section 2.4.2. and Section 2.4.3.) the test of statistical significance will not be applied due to insufficient size of the examined samples.

All χ^2 tests will be also based on 2x2 contingency tables and individual frequencies of investigated lexical items will be collapsed into one total value. Such a decision has been taken based on the presumption that all individual lexical items investigated in the present study represent a category used as a parameter in the operationalisation of a specific hypothesis.

2.6. Limitations of the applied methodology

The methodological assumptions presented in this Chapter show what procedures and guidelines will be followed in the forthcoming analysis of recurrent patterns in interpreting referred to in this study as universals.

The author of the present dissertation is fully aware of the controversy behind the term universals (see Chapter One), but decided to use it to refer to the investigated features of simplification, explicitness and normalisation since their ‘universality’ has been

⁴ Degrees of freedom = (rows - 1) x (columns - 1)

frequently hypothesised in the field of Translation Studies. The choice to investigate those features, commonly denoted as universals, on corpora of interpretations and translations from four different source languages was dictated by the will to validate or disprove the so much debated universality of their character.

Moreover, the analysis of interpreting and translation monolingual comparable corpora does allow to discern global patterns present in texts interpreted and translated from different languages, but makes it impossible to access the source texts and hence put forward viable explanations. The motivation behind adopting this approach was to provide a coherent analysis of interpretations and translations from four different languages, whereby the data set has been compiled according to unified criteria, since it rarely happens that interpretations and translations from four different languages are compared within one study. It is hypothesised that if the investigated features are universal they should be visible across all interpretations and translations from different source languages when these are compared against texts originally produced in the target language.

Moreover, it has proved to be impossible to achieve a perfect balance in all interpreting subcorpora of TIC with regard to the mode of delivery of the source speech. The subcorpus of interpretations from French is most heavily dominated by speeches that were originally read out, slightly less so is the subcorpus of interpretations from Dutch. The subcorpus of speeches interpreted from Spanish is balanced with this respect, while the source speeches of interpretations from German were predominantly spoken. This limitation will be taken into account in the analysis of the outcomes. Where possible, a control analysis will be carried on the subcorpus of interpretations from Spanish split into halves based on the mode of delivery of the source speech. Such an additional analysis will help to estimate to what extent the mode of delivery is likely to affect the end result. This is probably one of the most serious limitations of the study. On the other hand, this thesis is probably the first one that attempts to account for such differences at all e.g. in the analysis of simplification in interpreting. Earlier studies conducted on EPIC have not discussed the impact of this factor at all (see the discussion of Sandrelli – Bendazolli’s analysis of simplification in Section 1.5.1.).

Furthermore, as mentioned in Section 2.2.1.2. finding an ideal reference corpus for the translation subcorpora of TIC poses a methodological challenge. For the reasons spelled out in the referred section of this chapter, it has been decided that the translation corpora will be compared against speeches read out by native English Members of the European

Parliament. Such a choice was made since the examination of translated texts is here used as a reference to the analysis of interpreted discourse and no conclusions regarding interpreting universals that remain in the focal point of the present study are made based on this corpus. Still, it needs to be stressed that the results of such an analysis must be treated very cautiously.

Another limitation concerns the corpora of spoken interpreters' discourse that are subject to analysis in the case study. The recordings and their transcriptions have been generated for research purposes and the recorded interpreters' speeches are less formal than the speeches given at the European Parliament. This might have a bearing on the outcomes of the present analysis since formal register differs from the informal one.

It must also be stressed that the size of corpora comprising texts produced by individual interpreters are very small and insufficient to make generalisations. It is believed, however, that their size is appropriate to reveal the hypothesised universal tendencies (which as universal should be distinctive in every interpretation) or their conspicuous absence, as well as strong stylistic preferences.

While all the limitations presented above are important and should be addressed in suitable sections of the following chapters, they do not undermine the purpose of the study.

Chapter 3: Macro analysis of recurrent patterns in interpreting

3.1. Introduction

The motivation behind the macro analysis described in this Chapter is to examine whether simplification, explicitness and normalisation i.e. putative translation universals are also characteristic of interpreting, which is a mode of translation. For this purpose a series of analyses of different parameters assumed to be symptomatic of the tendency to simplification, explicitness and normalisation will be carried out in this Chapter.

First, to verify the simplification hypothesis all subcorpora will be compared with respect to the proportion of the corpus covered by list head, lexical density and the frequency of the most frequent English words. Secondly, the tendency to explicitness will be measured in the examination of the use of optional *that*, frequency of linking adverbials and apposition markers across all subcorpora. Finally, it will be analysed which of all respective subcorpora are characterised by most normalised language judging by the frequency of common lexical bundles and fixed expressions most frequently used in spoken English.

It is hypothesized that all interpretations are more simplified, explicit and use more normalised language than speeches originally produced in the target language.

3.1.1. Simplification

The first of the presumed translation universals discussed in this study has not been corroborated by the analysis of interpreting and spoken subcorpora of TIC. All three indicators presented in the tables below show how lexically diverse examined texts are. Laviosa's findings (1998: 8) implicate that when compared to texts originally produced in English translated texts show a tendency towards simplification i.e. that a larger proportion of the corpus is represented in the list head, lexical density is lower, while the proportion of most frequent English words is higher. These findings have, in general, not been confirmed by the analysis of interpreted discourse.

3.1.1.1. List heads

The percentage value expressing the proportion of the total sum of hundred most frequent words in the analysed corpus (list head) to the total number of running words reveals how repetitive the language in the examined text or corpus is. The higher the value, the bigger part of the corpus that the list head represents. In other words, a higher value points to the degree of repetitiveness. The latter, in turn, equals less variety and is in this study perceived as a tendency to simplification. Were Laviosa's (1998: 8) findings to be corroborated, the list heads of the interpreting subcorpora should represent a greater percentage of the respective corpora than in the case of native English speeches subcorpus. In other words, the values for SI_FR_EN, SI_ES_EN, SI_DE_EN, SI_NL_EN should be higher than the values observed in the subcorpus ORG_SP_EN.

As visible in Table 3, the number of tokens covered by the list head in ORG_SP_EN amounts to 29122, which constitutes 57.90 % of the entire corpus of 50295 words. The values observed in SI_FR_EN are lower, i.e. the number of tokens in the list head reach 28582, that is 56.84% of the corpus counting 50285 words in total. A 2x2 contingency table was used to calculate the disparity between both subcorpora with respect to the total number of tokens covered by the list head and the sum of the remaining ones by means of Chi-square test (see Table 3)⁵.

⁵ All remaining Chi-square tests in Chapter Three and Chapter Four are conducted analogically.

Table 3. The number of tokens in the list head vs. the number of remaining tokens in ORG_SP_EN and SI_FR_EN

	ORG_SP_EN	SI_FR_EN
Tokens in list head	29122	28582
Remaining tokens	21173	21703

The Chi-square calculated on the basis of Table 3 shows that the difference is statistically significant ($\chi^2=11.60$, $p<0.001$). Thus the language used in SI_FR_EN goes against the simplification hypothesis because it demonstrates significantly lower tendency to repetitiveness and can therefore be deemed more lexically diverse (less simplified with this respect).

On the other hand, looking at the percentage values calculated for all examined subcorpora presented in Table 4, one can infer that the language used in three out of four subcorpora, i.e. SI_ES_EN, SI_DE_EN, SI_NL_EN is more repetitive than the ORG_SP_EN.

Table 4. Proportion of the corpus covered by list head in TIC oral subcorpora (the higher the values, the more repetitive i.e. simplified the text).

	ORG_SP_EN	SI_FR_EN	SI_ES_EN	SI_DE_EN	SI_NL_EN
% Text covered by list head	57.90%	56.84%	58.88%	58.46%	58.15%

However, out of these three interpreting subcorpora only SI_ES_EN is statistically significantly different from ORG_SP_EN (see Table 5 below), while SI_DE_EN, SI_NL_EN and ORG_SP_EN are homogenous.

Table 5. Differences between oral corpora of TIC with respect to list head proportion: results of Chi-square tests calculated on raw counts in 2x2 contingency tables.

	ORG_SP_EN vs. SI_FR_EN	ORG_SP_EN vs. SI_ES_EN	ORG_SP_EN vs. SI_DE_EN	ORG_SP_EN vs. SI_NL_EN
χ^2	11.60	9.10	3.27	0.61
p-value	$p<0.001$	$p<0.01$	$p>0.05$	$p>0.05$

Analyzing simplification, one has to consider the mode of delivery of the source speeches dominant in respective interpreting subcorpora and its possible impact on the outcomes. A greater number of French and Dutch speeches that constitute the source of the interpretations included in SI_FR_EN and SI_NL_EN were read out, while in the case of SI_DE_EN the proportion was reverse and more source speeches were delivered orally in unscripted or semi-scripted way. SI_ES_EN is balanced with this respect (see Figure 9).

Assuming that the mode of delivery of the source speech is a decisive factor, it would be reasonable to expect that the discussed values regarding the proportion of list heads should descend in this order: SI_DE_EN, SI_ES_EN, SI_NL_EN, SI_FR_EN. Additionally, since SI_FR_EN and SI_NL_EN have a very similar proportion of source speeches that were spoken or read out, the values observed in these subcorpora should be similar. Neither of the two expectations have been confirmed. The descending order is different. The only subcorpus balanced with respect to the mode of delivery of the source speeches ranks the highest. The language used in SI_NL_EN and SI_FR_EN seems indeed to be least repetitive of all interpreting subcorpora, but they are significantly different from each other ($\chi^2=17.55$, $p<0.001$) and indeed SI_NL_EN resembles the two remaining interpreting subcorpora more than SI_FR_EN. Thus, it appears that the tendency to repetitiveness may be a result of an interplay of various factors rather than solely dependent on the mode of delivery of the original speech.

The mode of delivery, cannot be disregarded as, at least in Spanish, its impact on repetitiveness is significant. Since SI_ES_EN is the only balanced corpus with respect to the mode of delivery of the source speeches and guarantees a relatively larger number of speeches interpreted from the two modes of delivery, it has been split accordingly and analysed with respect to the percentage of text covered by list head. The outcomes are presented in Table 6 and Table 7 below.

Table 6. Proportion of the corpus covered by list head in the subcorpora of native English speeches, interpretations of Spanish speeches that were read out and interpretations of Spanish speeches delivered orally (the higher the values, the more repetitive i.e. simplified the text).

	ORG_SP_EN (24003 tokens)	SI_ES_EN_wr (24004 tokens)	SI_ES_EN_sp (23975 tokens)
% Text covered by list head	58.16%	58.65%	59.88%

Table 7. Statistical significance of differences in the proportion of the corpus covered by list head in the subcorpora of native English speeches, interpretations of Spanish speeches that were read out and interpretations of Spanish speeches delivered orally.

	ORG_SP_EN vs. SI_ES_EN_sp	ORG_SP_EN vs. SI_ES_EN_wr	SI_ES_EN_sp vs. SI_ES_EN_wr
χ^2	14.67	1.18	7.52
p-value	$p<0.001$	$p>0.05$	$p<0.01$

Native English speeches are least repetitive of all three subcorpora, but, in fact, they do not differ significantly from interpretations of speeches that were originally read out in Spanish. Interpretations of speeches that were originally spoken are significantly more

repetitive. This suggests that mode of delivery may have a significant impact on the degree of repetitiveness of the target text.

To what degree this tendency may be considered universal, it yet remains to be verified on other language pairs. For technical reasons conducting a reliable comparison between the other three source languages is impossible within the scope of the present study. Additionally, the corpora used in this particular examination are half as small as all the subcorpora used in the remaining tests in this Chapter. The observed results, however, undermine the assumption that the repetitiveness of simultaneous interpretations is independent from the mode of delivery of the original speech than non-interpreted discourse.

One of the possible explanations of such a result might be that interpreters rely on cognates more heavily than the translators. A study confirming such a trend in Hebrew-English language pair has been carried out by Shlesinger and Malkiel (2006: 8-10). Assuming that there are more cognates in French – English language pair than in other pairs represented in TIC, this might be a reason why the language of interpretation from French is less repetitive. The aforesaid assumption is actually plausible, as it has been also made in literature on contrastive linguistics e.g. Bailey – Maroldt (1977: 21–53) claim even that due to heavy lexical borrowing English and French are typologically close and that this relation may be stronger than the one between English and other Germanic languages. Additionally, in the EU context English and French constantly compete to play the role of the lingua franca at the EU institutions, which makes it more likely that many of the Eurojargon terms might be more similar in this language pair than in case of any other language combination.

Another confirmation of the above putative explanations can be found in the answer to one of the warming up questions provided by one of the interpreters, whose interpreting and speaking style is analysed in Chapter Four. When inquired, if there is a difference in interpreting from different languages he explicitly stated that in case of difficulties one could ‘transcode fairly easily word for word from French into English’, which would result in a less elegant, but decipherable interpretation. At the same time, he stated that the same procedure would not be possible in the interpretation from German. Given that interpreting at the plenary sittings of the European Parliament is considered to be very challenging, it is likely that the interpreters resort to such solutions.

A greater use of cognates and consequently also more awkward lexical formulations may increase lexical variety at the same time making the language less repetitive. This might be a reason why the language of interpretation from French in SI_FR_EN is less repetitive than interpretations from other source languages of TIC. A contrastive analysis of both source texts and the interpretations should be carried out on a parallel corpus to confirm such a hypothesis. Without the access to the source texts the above ideas remain only a speculation.

3.1.1.2. Lexical density

The next parameter used in this study in the investigation of simplification involves a comparison of lexical density of native English speeches and interpreted texts, which signifies how informative a text is. Lexical density is expressed as the proportion of lexical words to the total number of running words in the corpus. The confirmation of the simplification hypothesis (and Laviosa's findings) would require that lexical density in all interpreting subcorpora of TIC be lower than the lexical density of the subcorpus of native English speeches. Table 8 points to a completely opposite trend, which contradicts the simplification hypothesis and Table 9 shows that the differences between the values observed in the subcorpus of native English speeches and all interpreting subcorpora are statistically significant.

Table 8. Lexical density in TIC oral subcorpora (the lower the values, the less lexically dense i.e. simplified the text).

	ORG_SP_EN	SI_FR_EN	SI_ES_EN	SI_DE_EN	SI_NL_EN
Lexical density	54.46%	56.12%	55.67%	56.07%	56.31%

Table 9. Differences between oral corpora of TIC with respect to lexical density: results of Chi-square tests calculated on raw counts in 2x2 contingency tables.

	ORG_SP_EN vs. SI_FR_EN	ORG_SP_EN vs. SI_ES_EN	ORG_SP_EN vs. SI_DE_EN	ORG_SP_EN vs. SI_NL_EN
χ^2	27.81	14.85	26.35	34.75
p-value	p<0.001	p<0.001	p<0.001	p<0.001

Both tables point to a uniform and statistically significant trend. The language used in all interpreting subcorpora uses more lexical words than the language of native English

speeches. Lexical words carry greater meaning than function words, hence, more lexically dense language is also more informative.

Again, reference has to be made to the source speeches, the interpretations of which are included in the respective interpreting subcorpora. As written texts are usually more lexically dense than spoken ones, there is a reason to believe that subcorpora comprising a greater proportion of interpretations from speeches that were originally read out and not spoken would demonstrate greater lexical density. In such hypothetical scenario, it would be expected that the subcorpora comprising speeches interpreted from French and Dutch would be characterised by the highest lexical density, as opposed to the subcorpus of interpretations from German. The interpretations from Spanish would be placed in the middle of the continuum (for details regarding the prevailing mode of delivery of the original speeches in each subcorpus see Section 2.2.1.1. Figure 9). The actual results show that, while French and Dutch indeed demonstrate the highest lexical density, the values obtained in the analysis of the subcorpus of the interpretations from German are only slightly lower and interpretations from Spanish rank the last.

However, before it is assumed that the mode of delivery of the source speech does not have a key impact on the lexical density of the target text, it is worth examining how this parameter is reflected in the control examination of native English texts and the two sets of interpretations from Spanish involving two different modes of delivery of the original source speeches.

Table 10. Lexical density in the subcorpora of native English speeches, interpretations of Spanish speeches that were read out and interpretations of Spanish speeches delivered orally (the lower the values, the more simplified the text).

	ORG_SP_EN (24003 tokens)	SI_ES_EN_wr (24004 tokens)	SI_ES_EN_sp (23975 tokens)
Lexical density	54.23%	55.90%	55.45%

Table 11. Statistical significance of differences in lexical density in the subcorpora of native English speeches, interpretations of Spanish speeches that were read out and interpretations of Spanish speeches delivered orally.

	ORG_SP_EN vs. SI_ES_EN_sp	ORG_SP_EN vs. SI_ES_EN_wr	SI_ES_EN_sp vs. SI_ES_EN_wr
χ^2	7.29	13.64	0.98
p-value	p<0.01	p<0.001	p>0.05

Table 10 shows that native English speeches are less lexically dense than any interpretations from Spanish, regardless of the mode of delivery of the original speeches. Moreover, although at first glance the interpretations from unscripted oral speeches seem to

be slightly more simplified, there is no statistically significant difference between them and the interpretations of the speeches that were originally read out. Hence, it follows that the mode of delivery does not exert a significant impact on the lexical density of the interpreted text in this particular case. The tendency, has, again, been examined only on the corpora of Spanish and its universality should be tested by future research. As it is, however, this outcome lends support to the results presented in Table 8.

One of the possible causes of increased lexical density in interpreted texts may be interpreters' avoidance of redundancy and the need to produce a very compact and dense text caused by severe time constraint in simultaneous interpreting. Non-interpreted spoken discourse is characterised by a great tendency to redundancy and in standard situations speakers experience far less time pressure than interpreters that have to constantly keep up with the pace of the person they interpret. The interpreter, due to the time constraint has to condense the original message in order to deliver the same content in the minimum amount of time. Increased lexical density may, therefore, be a result of condensation techniques used by the interpreters to save time. The importance of condensation techniques and elimination of redundancies is emphasised also during interpreter training. This textual feature might therefore be attributed to the time-saving condensation techniques necessary in good interpretation and additionally reinforced by interpreter training.

Another tentative explanation may be linked to Schlesinger's (1995: 202) and Gumul's (2006: 182) observations regarding shifts from referential to lexical cohesion perceived as a form of explicitation. Both studies provide evidence suggesting that simultaneous interpreters may tend to substitute pro-forms or ellipsis used in the original text with lexical items, which are either repetitions or synonyms. Such a tendency, leading the interpreters to choose a lexical item rather than a pronoun or an auxiliary verb would also contribute to increased lexical density.

3.1.1.3. High frequency words

The final measure of simplification employed in this study helps to estimate the range of used vocabulary. The higher the proportion of the most frequent English words in the analysed corpus, the narrower the range of vocabulary i.e. the greater lexical simplification (for details of the investigated list of words see Section 2.3.1. and Appendix 2). The results

expected in line with the simplification hypothesis require all of the interpreting subcorpora to be characterised by a higher value, with regard to this parameter, than the subcorpus of native English speeches. The trend visible in Table 12 is completely reverse and Table 13 shows that the differences between the subcorpus of native English speeches and the respective interpreting subcorpora are statistically significant.

Table 12. The proportion of high frequency words in TIC oral subcorpora (the higher the values, the narrower the range of vocabulary i.e. simplified the text).

	ORG_SP_EN	SI_FR_EN	SI_ES_EN	SI_DE_EN	SI_NL_EN
% High frequency words	61.58%	57.71%	58.35%	59.62%	58.83%

Table 13. Differences between oral corpora of TIC with respect to the proportion of high frequency words: results of Chi-square tests calculated on raw counts in 2x2 contingency tables.

	ORG_SP_EN vs. SI_FR_EN	ORG_SP_EN vs. SI_ES_EN	ORG_SP_EN vs. SI_DE_EN	ORG_SP_EN vs. SI_NL_EN
χ^2	157.07	109.38	40.64	79.65
p-value	p<0.001	p<0.001	p<0.001	p<0.001

Hence, it seems likely that the range of vocabulary used in interpretation might be affected by the fact that the original speaker read out the speech, the results have to be analysed against the proportion of such speeches in respective interpreting subcorpora. Were the impact of this factor significant, judging by Figure 9 the most lexically sophisticated language would be found in SI_FR_EN and SI_NL_EN, while SI_ES_EN, and SI_DE_EN in particular, should manifest the greatest degree of simplification. The observed result indicate indeed that interpretations from French and interpretations from German are located on the opposite ends of the continuum, but surprisingly the lexis in interpretations from Dutch is less sophisticated than in case of the interpretations from Spanish. Based on this data, it is difficult to determine the real impact of the mode of delivery of the interpreted speech on lexical density, but a great contrast between SI_FR_EN and SI_DE_EN suggests that it is not an entirely irrelevant factor.

The examination of differences between the interpretations from speeches originally spoken and read out in Spanish does shed some light on this issue (see Table 14). Interpretations of speeches that were read out use significantly more sophisticated language than interpretations of originally unscripted speeches, which, on the other hand, are significantly less simplified than speeches originally given in English. The difference between interpretations of read speeches and spoken ones is statistically significant, but

also much smaller than between native English speeches and the interpretations of unscripted and semi-scripted speeches. This suggests that in case of interpretations from Spanish, the mode of delivery is an important factor, but the interpretations still use high frequency words less than native English MEPs.

Table 14. High frequency words in the subcorpora of native English speeches, interpretations of Spanish speeches that were read out and interpretations of Spanish speeches delivered orally (the higher the values, the more simplified the text).

	ORG_SP_EN (24003 tokens)	SI_ES_EN_wr (24004 tokens)	SI_ES_EN_sp (23975 tokens)
% High frequency words	61.71%	57.84%	59.06%

Table 15. Statistical significance of differences in the proportion of high frequency words in the subcorpora of native English speeches, interpretations of Spanish speeches that were read out and interpretations of Spanish speeches delivered orally.

	ORG_SP_EN vs. SI_ES_EN_sp	ORG_SP_EN vs. SI_ES_EN_wr	SI_ES_EN_sp vs. SI_ES_EN_wr
χ^2	35.14	74.86	7.41
p-value	p<0.001	p<0.001	p<0.01

It is very difficult to pinpoint exactly why interpreters tend to use less high frequency words. The most plausible explanation, already discussed in Section 3.1.1.2, seems to be the tendency to lexicalisation of referential cohesion observed by Schlesinger (1995: 202) and Gumul (2006: 182). Assuming that this was a universal tendency, the interpreters would lexicalise cohesive links expressed in the original in the form of referential cohesion. Consequently, in such situations, they would favour lexical forms to pronouns or auxiliary verbs and in the end use them less frequently. Hence pronouns and auxiliary verbs are included in the list of high frequency words, a tendency to lexicalise such items could explain the difference.

The difference in the proportion of high frequency words could be, in theory, caused by the use of synonyms, but data gathered in this study seems to contradict this assumption. If the interpreters used more synonyms, which could explain why they resort to high frequency words less, they would also be always less repetitive. As shown in Section 3.1.1.1. , interpretations from Spanish, German and Dutch are actually as repetitive as native English speeches. Hence, it is unlikely that the lower proportion of high frequency words is caused by greater use of synonyms.

3.1.1.4. Investigation of simplification in interpreting: Summary of the major findings

The tendency to lexical simplification in interpreting was investigated with respect to three key parameters: the percentage of the corpus represented by list head, lexical density and the proportion of high frequency words. On the whole, interpreting proves to be less lexically simplified than native English speeches, with one important exception i.e. repetitiveness.

Although the list head analysis of all interpreting subcorpora of TIC proved that they are as repetitive as the subcorpus of native English speeches (or even less so in case of interpretations from French), the more detailed examination of interpretations from Spanish suggests that the mode of delivery of the original speech may significantly influence this tendency as interpretations of unscripted and semi-scripted speeches are more repetitive i.e. simplified than native English speeches, while interpretations of speeches that were read out are significantly less repetitive i.e. not simplified at all. Yet, whether or not this tendency is universal or limited only to interpretations from Spanish into English needs to be verified by future research, as due to the structure of TIC it cannot be investigated in any more detail within the scope of the present study.

Examination of lexical density and the proportion of high frequency words showed, on the other hand, that interpreted discourse is by no means simplified with respect to these two parameters. The detailed examination of interpretations from Spanish suggest, that the mode of delivery of the source speech might have an influence on the level of lexical sophistication of the target text, but the tendency is not reversed.

3.1.2. Explicitness

The aim of this section is to verify, if explicitness could be a universal feature of the interpreted discourse. Hence the analysis focuses on the frequency of optional connective *that* after reporting verbs, occurrence of linking adverbials and the use of apposition markers. A higher frequency of the enumerated phenomena points to a greater tendency to explicitness. It is hypothesised that regardless of the source language interpretations are more explicit than native English speeches.

Because of a small frequency of the investigated phenomena in the oral subset of TIC, the analysis will not seek to verify if their distribution is equal in texts interpreted from speeches that were originally read out and spoken.

3.1.2.1. Optional *that*

Judging by the patterns of optional *that* revealed in the concordance lines of *admit, believe, claim, hope, know, say, suggest* and *tell* (in all forms) both translations and interpretations seem to be syntactically more explicit, in the sense that information is encoded in more overt way than in speeches originally produced in English (both spoken and read out). To what extent this tendency is to be ascribed to the process/ strategy of making implicit source text information explicit in target text, it is difficult to estimate. What is certain, however, is that interpretations from different source languages do display this tendency even when verbs supposedly influenced by interference are eliminated from the analysis. It is also striking that there is no statistically significant difference with this respect between interpretations from different source languages i.e. interpretations comprise a more or less uniform group that, on the other hand differs from speeches originally produced (spoken) in English.

As shown in Table 16 and Table 18 optional *that* is significantly more frequently verbalised rather than left out in all interpreted and translated subcorpora of TIC, regardless of the source language. In general, however, optional complementiser *that* is evidently more frequent in written language.

Table 16. Occurrences of optional *that* and zero-connectives after reporting verbs in TIC spoken subcorpora (z.c. stands for zero connective)

	ORG_SP_EN		SI_FR_EN		SI_ES_EN		SI_DE_EN		SI_NL_EN	
	that	z.c.	that	z.c.	that	z.c.	that	z.c.	that	z.c.
TOTAL	146	73	122	35	110	28	144	40	134	36
that %	67%		78%		80%		78%		79%	

Table 17. Differences between oral corpora of TIC with respect to occurrences of optional *that*: results of Chi-square tests calculated on raw counts in 2x2 contingency tables.

	ORG_SP_EN vs. SI_FR_EN	ORG_SP_EN vs. SI_ES_EN	ORG_SP_EN vs. SI_DE_EN	ORG_SP_EN vs. SI_NL_EN
χ^2	5.44	7.10	6.66	7.013
p-value	p<0.025	p<0.01	p<0.01	p<0.01

Table 18. Occurrences of optional *that* and zero-connectives in TIC written subcorpora, values normalised per 50000 and rounded (z.c. stands for zero connective)

	ORG_WR_EN		TR_FR_EN		TR_ES_EN		TR_DE_EN		TR_NL_EN	
	that	z.c.	that	z.c.	that	z.c.	that	z.c.	that	z.c.
TOTAL	142	39	242	9	199	17	234	22	144	17
that %	74%		96%		89%		91%		89%	

Table 19. Differences between oral corpora of TIC with respect to occurrences of optional *that*: results of Chi-square tests calculated on raw counts in 2x2 contingency tables.

	ORG_WR_EN vs. TR_FR_EN	ORG_WR_EN vs. TR_ES_EN	ORG_WR_EN vs. TR_DE_EN	ORG_WR_EN vs. TR_NL_EN
χ^2	33.78	11.74	17.52	9.51
p-value	p<0.001	p<0.001	p<0.001	p<0.01

Since optional *that* after the reporting verb *say* is particularly frequent both in the interpretations and the translations from German, there is a reason to believe that it could be caused by interference. In line with this assumption, the subcorpora of interpretations and of translations from German (SI_DE_EN and TR_DE_EN) have been compared with respective subcorpora of native English speeches (ORG_SP_EN and ORG_WR_EN) once again, the reporting verb *say* was excluded from the analysis. The trends have been confirmed i.e. the use of optional *that* is still more frequent in both interpretations and translations from German and the subcorpora remain statistically different from the respective subcorpora of native English speeches.

As has been demonstrated above, the tendency to use optional *that* connective is far more pronounced in interpreted and translated discourse than in speeches originally produced in English (both spoken and written). This tendency has been observed regardless of the source language and it needs to be stressed that interpreted texts are characterised by greater homogeneity. These results constitute a clear confirmation of the observations made by Olohan and Baker (2000) in their study conducted on Translational English Corpus, whereby optional *that* was significantly more frequent in texts translated into English from different languages.

One of the critical responses to the above mentioned study raised an important issue that should also be discussed here. Becher (2011: 31) observed that in the context of optional *that* one needs to know in how many of the source languages in the corpus the use of connective is optional and in how many it is compulsory. Consequently he infers that ‘the greater the number of source languages represented in TEC stipulating a complementizer after reporting verbs, the more likely it becomes that the higher occurrence

of reporting *that* in this corpus is the result of source language interference' (Becher 2011: 31). By analogy, this problem may also be vital for the present analysis, which revolves around two Romance and two Germanic languages that seem to treat complementizers differently: German uses optional complementizer *dass*, which can be used in reporting structures (Kenny 2005: 157); Dutch has a connective *dat* described in grammar books as obligatory (Donaldson 2008: 287, Fontein – Pescherter 2009: 187); French requires a complementizer, while Spanish favours it, but also allows for infinitive complementizer-free constructions (Becher 2011: 31). It seems therefore that the setup for analysis is ideal. This assumption, however, must be approached with great caution. Without the analyses of the source texts and the actual expressions that triggered the renditions with optional *that* in English, it remains impossible to determine with absolute certainty what is the factor contributing to the increased frequency of optional *that* in interpretations and translations. There are many variables that should be considered before one jumps into a far fetched conclusion that the presented results are a proof of explicitation only because in the examined set there are two source languages, in which the complementizer is optional and two that require its use and yet the frequency of optional *that* in their respective interpretations is similar. First, the fact that the complementizer is optional does not guarantee that it has not been used in all reporting structures in the source text (which would then mean that the starting conditions for interpreters from seemingly different languages are in fact the same). Secondly, the complementizer may be optional for some source language reporting verbs and obligatory for others and in different grammatical situations, which does not have to correspond to the requirements of the English reporting verbs tested in the present study. Thirdly, there is a chance that the prescriptive grammar does not fully reflect the use of the complementizer in a given language. Finally, it is possible that other grammatical structures in the source texts have been rendered in English with reporting verbs. In fact all four factors are hypothetically equally possible. House (2004: 187) suggests that the speakers of German prefer the 'aesthetics of spelling things out', which increases the possibility that, although the optional connective *dass* is not obligatory after reporting verbs, it may actually be relatively frequently used in such situations. On the other hand, although *dat* is compulsory in Dutch it has recently frequently been dropped in spoken discourse (Dynarowicz, personal communication). Moreover, Saldanha (2008: 22) claims that the closest Spanish equivalent to the English *say* and *tell* i.e. *decir* requires connective *que* in most cases, but there are

exceptions to that. Similarly, Kenny (2005: 167) provides also an example, where in her study a reporting structure followed by optional *that* in English was triggered by a completely different grammatical construction in the original German text. Of course whether or not a language requires or just allows optional complementizers after reporting verbs does remain an important factor, but it cannot be blindly assumed that the former case guarantees a high frequency of optional complementizers in the source text and the latter excludes it.

The examination of comparable corpora limits the analysis to the target text only and does not allow to investigate to what extent and which additional factors have affected the obtained results. It is also very difficult to control all possible variables in all examined languages. That is why a control corpus was used to aid the identification of those target language equivalents that could have been affected by source text interference. It was possible, because the interpretations and the translations originated from the same source texts. Therefore if optional *that* or zero connective was used excessively frequently in one of the corpora e.g. SI_EN_FR it could be compared with the other from this language pair e.g. TR_EN_FR. If both showed strikingly high frequency the verb was excluded from the analysis to minimise the effect of interference.

The reasons why interpreters and translators resort to optional *that* more frequently may be very diverse. Before analysing the results of their study Olohan and Baker (2000: 143-151) reviewed available literature on the use of *that* in English. In short, the verbalisation of optional complementizer *that* in English may depend on:

- the level of formality of the text,
- the matrix verb (some reporting verbs favour *that*),
- potential for ambiguity,
- adverbials used between the matrix verb and the clause (longer adverbials increase the possibility of *that*)
- cognitive complexity (explicit grammatical solutions are more frequent in cognitively complex environments).

Olohan and Baker (2000: 150) imply that increased frequency of optional *that* in translated texts may be potentially accounted for by the fact that translation is a relatively cognitively complex task. Outcomes presented in Table 16 and Table 18 above shed a new

light on such a hypothesis. For a number of reasons, the major one being the time constraint, simultaneous interpreting may be considered a cognitively more demanding task than translation. In line with this argumentation one should expect that interpreted texts demonstrate the highest frequency of optional complementizer of all analysed corpora, which is not the case. The frequency pattern, along which optional *that* dominates in the written counterpart of each set of subcorpora holds both for native English and translated/interpreted sets. In other words, when we compare any subcorpus from the oral set of TIC (e.g. SI_EN_NL or ORG_SP_EN) with its 'equivalent' in the written set (e.g. TR_EN_NL or ORG_WR_EN), the written corpus displays a greater frequency of optional complementizer *that*. When examining the use of optional *that* in interpreting, one has to take into account an interplay of at least three factors: interference, time constraint, modality and formality. Since the texts in the respective interpreting and translation subcorpora actually derive from the same source texts (in the same source language), the difference between them cannot be attributed to interference. Of course one could also argue that were it not for the time constraint, optional *that* could have been explicitated in simultaneous interpreting probably as frequently as in translation. Yet such an argument has to be refuted since the same time constraint allows simultaneous interpreters to resort to optional *that* on a greater number of occasions than native English Members of the European Parliament, whose time constraints are different. It is more likely that the change in modality and not the complexity of the task is the determining factor here, because such a change imposes different formality standards. Optional *that* would typically occur significantly more frequently in formal registers⁶, hence it is only logical to assume that the higher frequency rate points to a greater formality. Written texts are usually more formal than spoken ones which finds confirmation in distribution of optional *that* across all subcorpora of TIC.

Assuming that increased frequency of optional *that* observed both in interpreting and in translation compared with native English texts is in fact related to the use of more formal language one may also conclude that such a tendency supports the argument for a kind of hypercorrectness, language conservatism or in other words normalisation.

⁶ Biber (1995: 145) concludes that omission of *that*-complementizer is strictly related to informality. In his corpus *that* is omitted in all complement clauses in 85% in conversation, 60% in fiction, 30% in news and in 3% in academic prose.

Whatever the reason, it has been confirmed that, like translations, interpreted texts rely more heavily on optional complementizer *that*, which manifests a tendency to overtly encode syntactic information i.e. syntactic explicitness.

3.1.2.2. Frequency of linking adverbials

The next parameter used in this study to measure explicitness in texts across all subcorpora of TIC is the frequency of linking adverbials (listed in Section 2.3.2.). Table 20 demonstrates that interpretations from French, Spanish and Dutch as well as speeches originally produced in English are very similar with this respect. There is no observable pattern that could make the interpreted discourse more distinct from native English speeches. It is only the subcorpus of speeches interpreted from German that significantly stands out from other subcorpora, which, however, can mostly be ascribed to the use of one extremely frequent adverbial in that subcorpus i.e. *therefore*. Interestingly, the frequency of the same adverbial in translations from German is moderate.

Table 20. Occurrences of linking adverbials in TIC spoken subcorpora

	ORG_SP_EN	SI_FR_EN	SI_ES_EN	SI_DE_EN	SI_NL_EN
Linking adverbials	29	29	29	48	30

Table 21. Differences between oral corpora of TIC with respect to the occurrences of linking adverbials: results of Chi-square tests calculated on raw counts in 2x2 contingency tables.

	ORG_SP_EN vs. SI_FR_EN	ORG_SP_EN vs. SI_ES_EN	ORG_SP_EN vs. SI_DE_EN	ORG_SP_EN vs. SI_NL_EN
χ^2	0.00	0.00	4.70	0.02
p-value	p>0.05	p>0.05	p<0.05	p>0.05

Table 22. Occurrences of linking adverbials in TIC written subcorpora normalised per 50000 words.

	ORG_WR_EN	TR_FR_EN	TR_ES_EN	TR_DE_EN	TR_NL_EN
Linking adverbials	42	114	127	106	124

Table 23. Differences between oral corpora of TIC with respect to the occurrences of linking adverbials: results of Chi-square tests calculated on raw counts in 2x2 contingency tables.

	ORG_WR_EN vs. TR_FR_EN	ORG_WR_EN vs. TR_ES_EN	ORG_WR_EN vs. TR_DE_EN	ORG_WR_EN vs. TR_NL_EN
χ^2	35.32	45.29	29.27	42.86
p-value	p<0.001	p<0.001	p<0.001	p<0.001

Granted that a higher number of linking adverbials in a text demonstrates the author's greater tendency to explicitise, data in Table 22 and Table 23 confirm the hypothesis assuming that translations should be more explicit than speeches originally written in English. The frequency in the translation subcorpora is close to or more than three times higher as compared to texts originally produced in English.

Such a difference between interpreted language and translated language in the use of linking adverbials seems to be justified, since while interpreting simultaneously the interpreters cannot possibly envisage the structure of the entire speech. In other words, with this respect the interpreters may have to rely more on the authors of the original speech and their interference may be thus less visible. Translators, on the other hand do not face such severe time limitations and exercise a greater freedom in the use of linking adverbials. This problem pertains to what Shlesinger (1995: 194) refers to as 'the linearity constraint'. Both translators and interpreters are expected to apply local and global strategies to render cohesive ties successfully. What comes easily to translators, who enjoy the access to the full text at all times, is extremely difficult or impossible for interpreters. Not only the content of the speech is unveiled in fragments, but also the speaker's underlying intention may for long be unknown. In such a situation interpreters must be very cautious to minimise the possibility of misunderstanding.

It is particularly interesting to what extent interpretations and translations of, after all, the same source texts differ. The total frequencies of linking adverbials in respective translation corpora are, on average, four times higher in the case of translations from French, Spanish and Dutch. One possible interpretation of such distribution is that the original French, Spanish and Dutch speakers may be more or less as economical with linking adverbials as native English MEPs. Due to the linearity constraint part of those cohesive ties may be omitted in interpreting, which could be the reason why the texts are comparable in the number of linking adverbials to the native English speeches. Then the translators having the full access to the entire text render all linking adverbials, possibly also explicitating those cohesive ties that were rendered implicitly in the original speech. In the particular case of German, it is theoretically possible that the increased frequency of *therefore* in simultaneous interpreting corpus can be accounted for by a frequent use of its German equivalent *deshalb* in the source speeches. The linearity constraint may force interpreters to frequent straightforward rendition of the word, since their knowledge of the entire text, its structure and speaker's intentions is limited. In many cases such

straightforward renditions may, in fact, be the interpreter's last resort to preserve any of the cohesive ties of the source text. Translators from German, on the other hand, receive the full original speech and have the opportunity to work on the stylistics of the target text and find solutions eliminating the overrepresentation of *therefore*.

For the lack of access to the source text it is impossible to verify to what extent such a wild speculation is actually true. A further analysis on a set of parallel translation and interpreting corpora compiled of translations and interpretations of the same source texts could verify the above tentative explanations and bring other illuminating outcomes. For the present, however, such an analysis remains out of the scope of this study and may be only suggested for future research.

3.1.2.3. Frequency of apposition markers

The last parameter analysed to verify the explicitness hypothesis is the use of apposition markers *in other words, that is, that is to say, namely*. It is assumed that the greater the frequency of apposition markers in the corpus, the more lexically explicit the text.

As visible in Table 24, the interpreters use apposition markers more frequently than MEPs. The trend is most noticeable in interpretations from Dutch, French and Spanish, where apposition markers were used three times more frequently than in the speeches originally produced in English. The proportion of apposition markers in the interpretations from German outnumbers the one observed in original speeches twice.

Table 24. Occurrences of selected apposition markers in TIC spoken subcorpora

	ORG_SP_EN	SI_FR_EN	SI_ES_EN	SI_DE_EN	SI_NL_EN
TOTAL	5	15	15	13	14

Table 25. Differences between oral corpora of TIC with respect to the occurrences of apposition markers: results of Chi-square tests calculated on raw counts in 2x2 contingency tables.

	ORG_SP_EN vs. SI_FR_EN	ORG_SP_EN vs. SI_ES_EN	ORG_SP_EN vs. SI_DE_EN	ORG_SP_EN vs. SI_NL_EN
χ^2	5.00	5.00	3.56	4.27
p-value	p<0.025	p<0.025	p>0.05	p<0.05

Although there appears to be a consistent trend for the apposition markers to be more frequently used in interpreted discourse, the differences between only three of the interpreting subcorpora and the subcorpus of native English speeches are statistically

significant (see Table 25). The difference between native English speeches and the subcorpus of interpretations from German with respect to the frequency of apposition markers is close to statistical significance, but does not meet the prerequisite χ^2 value of 3.84 and thus exceeds the margin of error allowed in this study.

The difference between interpretations from Dutch and original speeches, which is statistically significant must have been affected by a strikingly high number of occurrences of the apposition marker *in other words* (11 out of 14) in the Dutch interpreting subcorpus. Interestingly the pattern is not reflected in Dutch translation subcorpus. A more detailed analysis revealed that the expression was used by five different interpreters, of whom two used it more often than others:

Table 26. Occurrences of *in other words* in Dutch interpreting corpus broken down by interpreters

Interpreter	Number of occurrences
B1	4
C1	1
M	1
T	2
U	3

Data in Table 26 suggests that the high frequency of *in other words* can in this case most probably be attributed to stylistic preferences of two interpreters. Although certainly interesting in depicting an observable trend, where interpreters tend to use more apposition markers than native English MEPs the data in this corpus does not make it possible for venturing any generalisations as it is skewed by the stylistic preference of a few interpreters. It appears that apposition markers are relatively infrequent, especially in spoken register of English and only testing a bigger sample may be more elucidating.

Table 27. Occurrences of selected apposition markers in TIC written subcorpora. Values normalised per 50000 words and rounded to the full digit.

	ORG_WR_EN	TR_FR_EN	TR_ES_EN	TR_DE_EN	TR_NL_EN
TOTAL	10	18	18	25	26

Table 28. Differences between written corpora of TIC with respect to the occurrences of apposition markers: results of Chi-square tests calculated on raw counts in 2x2 contingency tables.

	ORG_WR_EN vs. TR_FR_EN	ORG_WR_EN vs. TR_ES_EN	ORG_WR_EN vs. TR_DE_EN	ORG_WR_EN vs. TR_NL_EN
χ^2	35.31	45.29	29.27	42.86
p-value	p<0.001	p<0.001	p<0.001	p<0.001

The distribution of apposition markers does confirm that translations tend to be more explicit than the speeches originally written by native English MEPs. This trend is

strongly manifested in all translation corpora of TIC. The respective subcorpora are statistically significantly different from the speeches originally written in English. Although in case of translations the difference between the translations from German and native English speeches is statistically significant, it is also visible that the X2 value is much lower compared to the remaining values in Table 28 which reflects the trend visible in interpreting corpora. This suggests that to a certain extent the frequency of apposition markers might be influenced by the source language.

The above analysis shows that the observed trend confirmed initial assumptions and that in Translation and Interpreting Corpus apposition markers indeed prevail in interpreted and translated texts, when compared to native English speeches. The overall frequency of apposition markers in TIC is, however, too low to conclusively substantiate explicitness hypothesis. The above analysis shows that the frequency of occurrence of apposition markers should be compared between corpora more sizable than 50000 words.

3.1.3. Normalisation

The normalisation hypothesis stipulates that the language patterns characteristic of the target language are more frequently used in interpretation. The effect is supposed to be attributed to the interpreters' constant attempts to use normal (conventional) language resulting in exaggeration of the conventional linguistic patterns. In the present section the tendency to normalisation is investigated with two methods. The first involves a comparison of the frequency of repeated strings of words, i.e. lexical bundles common for all compared corpora, while the second focuses on fixed expressions most typical of a formal register of spoken English.

3.1.3.1. Lexical bundles (trigrams)

The first parameter used in the present study to verify the normalisation hypothesis involves the analysis of the frequency of lexical bundles (trigrams) i.e. recurring strings of words common for all subcorpora. The present analysis focuses on the comparison of the frequency of two lists of lexical bundles generated based on the reference subcorpora. The

reference subcorpus for the subcorpora of simultaneous interpretations is the subcorpus of native English speeches delivered orally (ORG_SP_EN), while translation subcorpora will be compared against a list generated from the subcorpus of native English speeches that were written and read out by native English MEPs (see Section 2.3.3.). Was the normalisation hypothesis to be corroborated, the listed lexical bundles should be more frequent in interpreting subcorpora than in subcorpora of native English speeches.

The list subject to examination in the oral subcorpora of TIC has been generated based on the subcorpus of unscripted and semi-scripted native English speeches according to the procedures described in Section 2.3.3. and includes the following lexical bundles: *and I think, at the moment, be able to, first of all, going to be, I think it, I think that, I think we, I would like, in terms of, is going to, make sure that, that we have, there is a, this is a, to have a, to make sure, we have to.*

Frequencies of certain bundles are higher in specific interpreting subcorpora: *and I think* in SI_DE_EN, *be able to* in SI_FR_EN, *I think that* in SI_DE_EN and SI_NL_EN, *to make sure, we have to* and *we need to* in SI_ES_EN. A comparison with the frequency of the enumerated bundles in the translation corpora helps to estimate the likelihood of such a high occurrence being a result of interference. Since the lexical bundles examined on oral and written TIC subcorpora have been developed based on two different reference subcorpora for each respective group (ORG_SP_EN for oral subcorpora and ORG_WR_EN for written) frequencies of only selected bundles can be compared across different modalities while others cannot. Trigram *we need to* is more frequent also in SI_ES_EN and in TR_ES_EN, which leads to the conclusion that the higher use of this lexical bundle might be caused by interference. Bundles like *and I think, I think that, to make sure* were not frequent enough to be analysed in written subcorpora of TIC, which makes it difficult to determine to what extent they might have been caused by interference and they were therefore not included in the repeated statistical test. On the other hand, *be able to* and *we have to* are relatively evenly distributed in translation subcorpora, which suggests that their higher frequencies in interpreting subcorpora may be less strongly affected by interference and they were not deleted from the repeated test. Phrases that could have been caused by interference have been removed from individual comparison. Hence, the comparison between ORG_SP_EN and SI_ES_EN was carried out excluding *to make sure* and *we need to, and I think* and *I think that* were not included in the repeated test of

ORG_SP_EN and SI_DE_EN, while *I think that* alone was not taken into account in the repeated examination of ORG_SP_EN and SI_NL_EN.

It seems that the deleted bundles have not skewed previously obtained results. All interpreting subcorpora tested once again proved to be significantly different from the corpus of native English speeches ($p < 0.001$). The test of lexical bundles has corroborated the normalisation hypothesis.

Table 29. Lexical bundles in TIC spoken subcorpora.

	ORG_SP_EN	SI_FR_EN	SI_ES_EN	SI_DE_EN	SI_NL_EN
TOTAL	476	547	716	627	622

Table 30. Differences between oral corpora of TIC with respect to the occurrences of lexical bundles: results of Chi-square tests calculated on raw counts in 2x2 contingency tables.

	ORG_SP_EN vs. SI_FR_EN	ORG_SP_EN vs. SI_ES_EN	ORG_SP_EN vs. SI_DE_EN	ORG_SP_EN vs. SI_NL_EN
χ^2	4.99	48.89	20.93	19.64
p-value	$p < 0.05$	$p < 0.001$	$p < 0.001$	$p < 0.001$

As visible in Table 29 and Table 30 the analysis proves that even patterns most characteristic of native English speeches occur far more frequently in subcorpora of interpreted English. The general trend can be observed regardless of the source language. It is, however, also visible that the difference in the frequency of the examined lexical bundles between the interpretations from French and native English speeches is far less pronounced than in the case of speeches interpreted from Spanish. This result remains in line with the finding that interpretations from French are also less repetitive and use fewer high frequency words than texts interpreted from other languages (see Section 3.1.1.). From the investigation of lexical bundles it follows that the language used in interpretation from French is less conventional than the language used in interpretations from other source languages but more conventional than the one of native English speeches. As it has already been suggested in Section 3.1.1. such an outcome might be an effect of greater lexical similarity of French and English, allowing the interpreters to ‘transcode’ word for word in difficult situations, leading to a higher use of cognates in interpretations from French, which may ultimately contribute to greater lexical variety and less conventional language.

The result may also potentially be affected by different mode of delivery of the original speeches (see Figure 9). The majority of French speakers, whose speeches constitute the original texts of the interpretations included in SI_FR_EN have read the text.

It is therefore possible that certain features of written register were transferred to the interpretations, which could be another viable reason why the language of interpretations from French is less conventional. Although the mode of delivery may exert some impact on the language used in the target text it seems not to be the key factor, as a similar effect would be then expected in the interpretations from Dutch that originate also mostly from speeches that were read out at the plenary. Yet, the result of the speeches interpreted from Dutch resembles more the one yielded in the analysis of interpretations from German, a greater proportion of which originates from speeches delivered orally. It seems, therefore, more likely that this parameter is more susceptible to the source language than the mode of delivery of the original speech.

Among the translation subcorpora of TIC, it is the subcorpus of speeches translated from Dutch that does not significantly differ from speeches originally written in English (see Table 31 and Table 32 below). Excluding from the analysis the bundles that may potentially skewer the outcome due to high frequencies in particular translation subcorpora (*in order to, the fact that, we need to*) does not change this result. A similar, albeit not so significant trend is also observable in translations from French, which, although statistically significantly different from speeches originally written in English still resemble them more than the translations from Spanish and German. Such results suggest that in case of translations the mode of delivery of the original speech may be more important for this parameter. As explained in Section 2.2.1.2. the reference corpus of speeches originally written in English and read out at the European Parliament is probably most ‘written’ in the entire set of TIC subcorpora. In other words, speeches that were written, even when meant to be read out still bear very strong characteristics of a written text. The vast majority of translations from French and from Dutch included in TIC translation subcorpora is based on texts that were originally written and read out in the source language (see Figure 9). These subcorpora are therefore naturally more characteristic of the written register than subcorpora of translations from Spanish and German, where the impact of spoken word was greater. It seems that in case of translation the mode of delivery is a more decisive factor than in case of interpretation.

Both interpretations and translations, however, are clearly also affected by interference, which is probably the reason why the frequency of lexical bundles is most frequent in translation from Spanish, even though the proportion of spoken discourse among original Spanish speeches was considerably greater than in the subcorpus of

translations from German. Similarly, although it is the subcorpus of translations from French that contains the greatest proportion of written texts among the source speeches, the Dutch subcorpus is statistically most similar to the speeches originally written in English. It follows that translation is strongly affected by both interference and mode of delivery of the target text, but the latter seems to have a greater impact in translation than on interpreting.

Table 31. Lexical bundles in TIC written subcorpora. Frequencies are normalised per 50000 words and rounded to the full digits.

	ORG_WR_EN	TR_FR_EN	TR_ES_EN	TR_DE_EN	TR_NL_EN
TOTAL	305	366	516	505	316

Table 32. Differences between written corpora of TIC with respect to the occurrences of lexical bundles: results of Chi-square tests calculated on raw counts in 2x2 contingency tables.

	ORG_WR_EN vs. TR_FR_EN	ORG_WR_EN vs. TR_ES_EN	ORG_WR_EN vs. TR_DE_EN	ORG_WR_EN vs. TR_NL_EN
χ^2	5.99	57.75	53.38	0.20
p-value	p<0.025	p<0.001	p<0.001	p>0.05

3.1.3.2. Fixed expressions

According to the normalisation hypothesis, patterns typical of the target language might be exaggerated in interpreted and translated texts. It was therefore assumed that most frequent fixed expressions typical for professional English i.e. *as far as*, *as to*, *as well*, *as well as*, *at all*, *first of all*, *in fact*, *in favour of*, *in order to*⁷, *in terms of*, *kind of*, *of course*, *on behalf of*, *once again*, *sort of*, *with regard to* should be better represented in translation and interpreting corpora. Since the set of fixed expressions has been proved to be most characteristic of spoken English it was also anticipated that they will be less frequent in written translations of the speeches. Similarly, since speeches contained in ORG_WR_EN have been written and read out i.e. are more literate than oral, it is expected that they will occur even less frequently in this subcorpus (see Figure 11).

The first hypothesis has not been confirmed. The queried fixed expressions are used significantly more frequently only in interpretations from French and from Spanish while interpretations from German and from Dutch seem not to differ significantly from the

⁷ Expressions *in order to* and *on behalf of* have been examined both in Section 3.1.3.1. and Section 3.1.3.2. as they have been generated by two independent methods and for that reason it has been decided that they should not be eliminated from either investigation.

native English speeches. As shown in Table 33 the frequency of investigated expressions is indeed higher, but the upward trend is not strong enough (see Table 34). It seems therefore that this particular parameter is very much dependent on the source language and that in this case interpretations from Romance languages are characteristic of a higher frequency of fixed expressions thus showing a greater tendency towards conventional language.

Table 33. Occurrences of the most frequent fixed expressions in TIC spoken subcorpora.

	ORG_SP_EN	SI_FR_EN	SI_ES_EN	SI_DE_EN	SI_NL_EN
TOTAL	225	277	270	263	259

Table 34. Differences between oral corpora of TIC with respect to the occurrences of most frequent fixed expressions: results of Chi-square tests calculated on raw counts in 2x2 contingency tables.

	ORG_SP_EN vs. SI_FR_EN	ORG_SP_EN vs. SI_ES_EN	ORG_SP_EN vs. SI_DE_EN	ORG_SP_EN vs. SI_NL_EN
χ^2	5.42	4.11	2.98	2.40
p-value	p<0.025	p<0.05	p>0.05	p>0.05

As assumed the investigated expressions are more frequent in translation subcorpora than the subcorpus of original speeches written and read out in English. However, the results regarding the written subset of TIC have to be approached with great caution. On the one hand, it may be inferred that translators use fixed expressions more frequently, on the other, the fact that original native English speeches have been written might additionally strengthen the effect since the list of tested expressions was developed for spoken texts.

Table 35. Occurrences of the most frequent fixed expressions in TIC written subcorpora. Frequencies are normalised per 50000 words and rounded to the full digits.

	ORG_WR_EN	TR_FR_EN	TR_ES_EN	TR_DE_EN	TR_NL_EN
TOTAL	137	202	202	244	218

Table 36 Differences between written corpora of TIC with respect to the occurrences of most frequent fixed expressions: results of Chi-square tests calculated on raw counts in 2x2 contingency tables.

	ORG_WR_EN vs. TR_FR_EN	ORG_WR_EN vs. TR_ES_EN	ORG_WR_EN vs. TR_DE_EN	ORG_WR_EN vs. TR_NL_EN
χ^2	13.15	12.93	32.33	19.46
p-value	p<0.001	p<0.001	p<0.001	p<0.001

As in the case of optional *that* connective, several values are higher, which might suggest a greater influence of language transfer or source text interference. Expressions *as far as* and *in fact* are overrepresented both in interpretations and translations and therefore a

control analysis has been carried out excluding the two expressions. The repeated test has, however, not affected the observed trends.

As demonstrated above, contrary to the results obtained in the previous section, the normalisation hypothesis has not been corroborated. Fixed expressions most frequent in spoken professional English are better represented in the interpretations from Romance languages thus proving that their frequency is dependent on the source language. The latter observation apparently confirming normalisation hypothesis in written subcorpora of TIC cannot be treated as the ultimate corroboration of the normalisation hypothesis for translation since it is very plausible that the outcome might have been affected by the fact that the features of orality might be less pronounced in the reference corpus of written English speeches.

3.2. Observed patterns, their universality and idiosyncratic stylistic variation

The overriding aim of the analysis presented in this Chapter was to establish if the features by many believed to be translation universals i.e. simplification, explicitness and normalisation are characteristic of interpreting and to what extent they are subject to idiosyncratic stylistic variation and impact of the source language.

A wide range of parameters have been tested to verify the above hypothesis. These included the examination of the proportion of the corpus covered by list head, lexical density and the frequency of the most frequent English words applied in earlier studies to test the tendency of the translated texts to simplification. Furthermore, parameters such as the use of optional *that*, frequency of linking adverbials and apposition markers have been used to see if interpreted texts are more explicit than speeches originally produced by native English speakers. Finally, the language of all respective subcorpora has been tested for normalisation, which involved the analysis of the frequency of common lexical bundles and fixed expressions most frequently used in spoken English.

The null hypothesis assumed that all interpretations are more simplified, explicit and use more normalised language than native English speeches and that the same pattern is visible in the interpretations produced by individual interpreters but is modified by idiosyncratic stylistic variation to a certain extent.

It has been shown that no general pattern pointing to a more simplified language can be observed in interpreting subcorpora of TIC. The analysis of the list heads of oral subcorpora of TIC proves that the language used in interpretations from Dutch and German does not differ significantly from native English speeches in this respect. On the other hand, Spanish appears to be significantly more and French significantly less repetitive. Such a differentiation across interpretation from different languages shows that there is no universal pattern that could suggest that interpreted language is in general more repetitive and hence simplified than non-interpreted discourse. The remaining two parameters point more consistently to the fact that, if any global pattern could be observed, it is completely opposite to simplification. In the analysed oral subcorpora of TIC native English speeches proved to be significantly more simplified with respect to lexical density and the percentage of high frequency words. If any generalisation at all was to be made based on this outcome, it could be argued that interpreted language tends to be more lexically dense and more sophisticated in terms of vocabulary.

A tendency to explicitness has been conclusively confirmed by only one out of three tested parameters, while another appears to indicate an observable trend that has not been statistically significant. It has been shown that similarly to translations and contrary to the native English speeches, interpretations in TIC are characteristic of a higher frequency of optional connective *that* following reporting verbs, thus manifesting a tendency to syntactic explicitness. A similar, though not statistically significant trend may be observed in the distribution of apposition markers across oral subcorpora of TIC. The interpretations rely more heavily on apposition markers, yet the linguistic phenomena themselves seem to be used relatively rarely in general, and in spoken discourse in particular, and a far more sizeable corpus of interpretation would be necessary to conclusively prove or disprove the tentative assumption that interpretations as such are more abundant with apposition markers than speeches originally produced in English. Finally the occurrence of linking adverbials does not point to any general patterns in interpreting. Except for the interpretations from German, where one word i.e. *therefore* is used excessively, the frequency of linking adverbials in all oral subcorpora of TIC is very consistent and the remaining subcorpora of simultaneous interpretation do not differ significantly from native English speeches.

The first parameter applied in this study to point to patterns indicating a tendency to normalisation yields conclusive results and corroborates the hypothesis. The frequency of common lexical bundles does suggest that trigrams common for all oral subcorpora of TIC

occur in a greater number in interpreting subcorpora, and even after a few doubtful cases were excluded from the analysis the trend is still statistically significant. The second parameter testing the distribution of the most frequent fixed expressions across oral subcorpora of TIC suggests that only when working from Romance languages do the interpreters make the greatest use of such most frequent phrases thus proving that their language of interpretations shows a tendency to normalisation with this respect. Thus the outcomes of the analysis of the second investigated parameter do not corroborate normalisation hypothesis.

In the final analysis none of the hypothesised universals has proved to be indeed a universal feature that could be observed in all interpreting subcorpora. Different parameters point to different trends or no trends at all. Although no feature has proved to be ultimately universal, certain global patterns have been observed but since they are not discernible across all interpretations it would be difficult to presume that their origin lays solely in the nature of the interpreting process.

Table 37. Juxtaposition of the outcomes of macro analysis.

Examined universal	Tested parameters	Has the universal been confirmed?
SIMPLIFICATION	List head	no
	Lexical density	no (interpreted texts are less simplified)
	High frequency words	no (interpreted texts are less simplified)
EXPLICITNESS	Optional connective <i>that</i>	yes (interpreted texts are more explicit)
	Linking adverbials	no
	Apposition markers	yes* (interpreted texts are more explicit)
NORMALISATION	Lexical bundles	yes (interpreted texts are more normalised)
	Fixed phrases	no

* denotes an observable, but not statistically significant trend.

Chapter 4: Micro analysis of idiosyncratic variation: interpreting style

4.1. Introduction

The present Chapter reports on the part of the study devoted to micro analysis that was conducted in an attempt to provide answers to three out of four of the research questions posed in Chapter Two:

- (2) Are individual interpreters consistent in their tendencies to simplify, explicitise and normalise regardless of the source language?
- (3) Do individual interpreters differ from each other in their tendencies to simplification, explicitness and normalisation?
- (4) Is an individual's interpreting style similar to their speaking style?

To achieve this goal a set of linguistic features revealing a tendency to simplification, explicitness and normalisation will be investigated following the method described in Section 2.4. The results of the corpus-based study will be analysed against the background of non-linguistic information elicited from the examined interpreters prior to the recording of their non-interpreted discourse.

4.2. Non-linguistic data from introspection

The two interpreters, whose style is investigated in the present Chapter are native British English speakers awarded with a diploma of higher education, who have interpreted for the European institutions for at least fifteen years. They are both male and have lived outside Britain for many years now. In the warm-up session leading to the actual recording of their non-interpreted discourse the two interpreters were asked introductory questions about their education, career, nature of the interpreting job at the EU institutions and nature of interpreting itself. Responses to all the questions cannot be revealed, as they include personal details and the interpreters expressly wished to remain anonymous. Replies to two questions may, however, shed light on the interpretation of the data presented in this Chapter. The answers have been summarised in Table 38.

Table 38. Summary of the answers given by interpreters in the warm-up session.

Questions:	Interpreter I1	Interpreter I2
Can you see any difference while interpreting from different languages?	- there is no difference with respect to the structure of the original language, but one feels less confident in interpretation from languages they acquire last (in this case Spanish)	- in difficult cases one can transcode word for word from French into English, and the interpretation will be still comprehensible (which is impossible in interpretation from German)
Does the language you speak differ from the language of your interpretation?	- the language of interpretation is determined by the style of the original speaker, which the interpreter is obliged to follow	- interpretation involves higher register, formal language and greater use of technical terminology

The interpreters present slightly different approaches in answering both questions and it seems that their attitudes are also partially reflected in the way they interpret. In the light of the above presented answers, one could expect that linguistic features of I1's interpreted discourse would be far less consistent, if indeed his priority was to mimic the style of the original speakers. Had the level of interpreter's confidence any impact on interpreting style, then probably I1's interpretation from Spanish could stand out from the interpretation from two other languages. In case of I2, one may expect that were he indeed occasionally to transcode from French word for word, the language of this interpretation could be less repetitive and perhaps use more unconventional lexis. It would also be expected that he would be more formal in his interpretation.

4.2.1. Simplification

The tendency to simplification has been measured following the parameters described in Chapter Two and applied already in the analysis of recurring patterns of interpreting (see Section 3.1.1.). These are: the proportion of the corpus covered by list head indicating the degree of repetitiveness, lexical density pointing to lower or higher informativeness, and the percentage of high frequency words signifying the level of lexical sophistication.

Due to a great susceptibility of all three parameters to corpus size, the full collections of texts produced by the two interpreters, whose style is subject to analysis, could not have been used. That is why it has been decided that simplification will be investigated on corpora comprising carefully selected text samples. Each interpreting corpus tested in Section 4.2.1. consists of samples of four interpretations delivered by the interpreters I1 and I2 of speeches that were read out at the European Parliament by four different source speakers. Samples include 60-word fragments from the beginning, ending and middle sections of the text. Thus each sample is 180 words long and each corpus counts 720 words.

To increase the comparability of the analysed texts, samples were selected from interpretations of speeches delivered in only one mode i.e. speeches that were originally read out. Unfortunately, an analogous comparison of interpretations of unscripted and semi-scripted speeches is not possible due to the lack of data.

4.2.1.1. List heads

The examination of list heads allows to compare analysed corpora with respect to lexical repetitiveness, which is one of the parameters of simplification examined both in the macro and micro analysis of the present study. Data in Table 39 and Table 41 shows the percentage value standing for the proportion of the total sum of hundred most frequent words in the analysed corpus (list head) to the total number of running words. The higher the values, the bigger part of the corpus that the list head represents. Consequently, a higher value points to a greater degree of repetitiveness i.e. less variety and a tendency to simplification.

Table 39. Proportion of the corpus covered by list head analysed on 720 word corpora of interpreter I1's non-interpreted and interpreted discourse consisting of text samples.

	I1_ORG_SP_EN	I1_SI_FR_EN	I1_SI_ES_EN	I1_SI_DE_EN
% Text covered by list head	68%	66%	70%	73%

Table 40. Repetitiveness of I1's interpreted and non-interpreted discourse: Results of Chi-square tests calculated on raw counts in 2x2 contingency tables.

	I1_ORG_SP_EN vs. I1_SI_FR_EN	I1_ORG_SP_EN vs. I1_SI_ES_EN	I1_ORG_SP_EN vs. I1_SI_DE_EN
χ^2	1.02	0.38	3.21
p-value	p>0.05	p>0.05	p>0.05

The set of data presented in Table 39 seems to indicate that the language of interpreter I1 is most repetitive while he interprets from German, slightly less so, when the source language is Spanish, even less while he speaks on his own and the least when delivering an interpretation from French. Thus, no homogenous pattern valid for all I1's interpretations is revealed. The results, however, very much reflect the general trend observed in oral subcorpora of TIC (see Section 3.1.1.), where speeches interpreted from Spanish and German demonstrate a greater degree of repetitiveness than native English speeches and interpretations from French. The case of interpretations from French is particularly interesting, because also in the investigation of list heads in TIC did the interpretations from French manifest a markedly lower tendency to repetitiveness. In TIC, however, each interpreting subcorpus is characterised by a different proportion of the original speeches delivered orally and read out, whereas all interpretations delivered by interpreter I1 originate in speeches that were read out.

In case of the micro study of I1's interpreting style, however, the results obtained in the analysis of the respective interpreting corpora do not differ significantly from the outcome calculated for the subcorpus of I1's spoken discourse. The question is, whether the observed tendencies would hold true for a bigger and more representative corpora of I1's interpretations. From the statistical analysis of the present, very small samples (see Table 40), it can be inferred, that I1's speaking style and interpreting style do not differ significantly with respect to repetitiveness. This means that the I1's speaking style has a great impact on his interpreting and that he has not developed a uniform interpreting style that would be consistent in the interpretations from each source language and significantly different from his non-interpreted discourse. On the contrary, the interpretations from French and from German can be found on the opposite ends of the continuum, with I1's

non-interpreted discourse in between. While both are similar to the non-interpreted speech, they do deviate significantly from each other ($\chi^2 = 7.83$, $p < 0.01$), which indicates that I1's tendency to be more repetitive may be susceptible to the source text interference. These findings are in line with his belief that the language of interpretation is very much dictated by the original speaker's speaking style.

A completely different tendency can be observed in case of interpreter I2, whose interpreting style is significantly different with respect to repetitiveness than his speaking style. Regardless of the source language interpreter I2 uses a significantly less repetitive language in simultaneous interpreting than in non-interpreted discourse. His interpretations from French and German are highly homogenous ($\chi^2 = 0.11$, $p > 0.05$), which implies that when it comes to repetitiveness, his interpretation is 'resistant' to source language interference, or interference from either source language affects him in a very similar way. This is actually surprising, considering that I2 admitted that it is possible to transcode from French. It implies that either he does not, in fact, resort to this option very often or that it does not significantly affect the tendency to repetitiveness in his interpretation.

Table 41. Proportion of the corpus covered by list head analysed on 720 word corpora of interpreter I2's non-interpreted and interpreted discourse consisting of text samples.

	I2_ORG_SP_EN	I2_SI_FR_EN	I2_SI_DE_EN
% Text covered by list head	74%	67%	66%

Table 42. Repetitiveness of I2's interpreted and non-interpreted discourse: Results of Chi-square tests calculated on raw counts in 2x2 contingency tables.

	I2_ORG_SP_EN vs. I2_SI_FR_EN	I2_ORG_SP_EN vs. I2_SI_DE_EN
χ^2	9.06	11.18
p-value	$p < 0.01$	$p < 0.001$

It is highly significant, however, that the tendencies observed in I2's speaking and interpreting styles completely contradict the observations made on TIC, where interpretations from German are as repetitive, as the originals, while texts interpreted from French are significantly less repetitive. I2's results significantly undermine the hypothesised universality of patterns observed in TIC and point to the fact that the tendency to repetitiveness hinges more on stylistic idiosyncrasies and possibly on one's 'immunity' to interference.

It is also interesting to compare both interpreters to one another. Table 43 depicts statistical differences with respect to the tendency to repetitiveness between each interpreter's output.

Table 43. Repetitiveness of I1 vs. I2: Results of Chi-square tests calculated on raw counts in 2x2 contingency tables.

	I1_ORG_SP_EN vs. I2_ORG_SP_EN	I1_SI_FR_EN vs. I2_SI_FR_EN	I1_SI_DE_EN vs. I2_SI_DE_EN
χ^2	6.28	0.028	5.28
p-value	p<0.025	p>0.05	p<0.025

It reveals that while in non-interpreted discourse I2 manifests a greater tendency to repetition than I1, there is no significant difference in the way both men interpret from French and that I1's language in interpretations from German is more repetitive. What follows is that regardless of their natural tendency to repetitiveness manifested in non-interpreted discourse, both interpreters are equally repetitive when working from French into English. Such a finding indicates that indeed the pattern regarding repetitiveness in interpretations from French observed in TIC might be a universal one and that the tendency to repetitiveness, or the lack of it, is heavily contingent on the source language and subject to interference. At the same time, the statistically significant difference between I1's and I2's interpretation from German suggests that there is a high chance that one's interpreting style might be consistent enough to be less subject to this influence.

4.2.1.2. Lexical density

The next parameter tested in the investigation of simplification is lexical density, which helps to establish how informative a text is. Lexical density is expressed as the proportion of lexical words to the total number of running words in the corpus. A lower proportion of lexical words to function words i.e. the lower lexical density suggests that the text is less informative and hence more simplified.

The superficial analysis of lexical density in Table 44 expressed in percentage suggests that I1's interpretations from Romance languages are less informative than interpretations from German and his non-interpreted discourse. As shown in Table 45, in fact, his output is very homogenous with respect to lexical density i.e. there is no statistically significant difference between the four corpora.

Table 44. Lexical density analysed on 720 word corpora of interpreter I1's non-interpreted and interpreted discourse consisting of text samples.

	I1_ORG_SP_EN	I1_SI_FR_EN	I1_SI_ES_EN	I1_SI_DE_EN
Lexical density	57%	56%	55%	57%

Table 45. Lexical density of I1: Results of Chi-square tests calculated on raw counts in 2x2 contingency tables.

	I1_ORG_SP_EN vs. I1_SI_FR_EN	I1_ORG_SP_EN vs. I1_SI_ES_EN	I1_ORG_SP_EN vs. I1_SI_DE_EN
χ^2	0.10	0.28	0.04
p-value	p>0.05	p>0.05	p>0.05

Such results are very surprising when compared to the general tendencies observed in TIC, where all interpretations regardless of the source language are statistically significantly different from native English speeches and demonstrate greater lexical density. It is the more unexpected, that in this section the source speeches of all interpretations were read out, which means that they all could have been influenced by the written word, while I1's non-interpreted discourse was not. Table 46 and Table 47 indicate that the investigation of lexical density in I2's interpretations and non-interpreted discourse reveals the same uniformity of all analysed texts as in the case of I1.

Table 46. Lexical density analysed on 720 word corpora of interpreter I2's non-interpreted and interpreted discourse consisting of text samples.

	I2_ORG_SP_EN	I2_SI_FR_EN	I2_SI_DE_EN
Lexical density	56%	55%	57%

Table 47. Lexical density of I2: Results of Chi-square tests calculated on raw counts in 2x2 contingency tables.

	I2_ORG_SP_EN vs. I2_SI_FR_EN	I2_ORG_SP_EN vs. I2_SI_DE_EN
χ^2	0.04	0.23
p-value	p>0.05	p>0.05

The most likely explanation of such findings is that if there are any differences between the analysed corpora of I1 and I2, then possibly the sample subject to examination in this section is too small to show it, which, however, suggests that they are not pronounced.

4.2.1.3. High frequency words

Finally, the last measure of simplification allows to determine the range of vocabulary used in the texts. A higher proportion of the most frequent English words in the analysed corpus

indicates a narrower the range of vocabulary i.e. greater lexical simplification (for details of the investigated list of words see Section 2.3.1. and Appendix 2).

As it can be seen in Table 48 and Table 49, but for interpretations from French, interpreter I1's interpreted and non-interpreted discourse is rather homogenous as regards the range of used vocabulary. There is no statistically significant difference between I1's non-interpreted discourse and the interpretations from Spanish and German. Interpretations from French are statically significantly more lexically sophisticated, not only than the non-interpreted texts, but also than interpretations from German ($\chi^2=6.58$, $p<0.025$). Thus, I1's interpretations from French and German seem to be statistically significantly different with respect to two features: repetitiveness and range of vocabulary and in both cases interpretations from French are more lexically complex. Such an outcome once again certifies to the fact that I1's interpretations from French are more subject to interference. There are fewer high frequency words in I1's interpretations from Spanish than in his interpretations from German and non-interpreted discourse. The number, however, slightly exceeds the values observed in interpretations from French. It is difficult to conclusively point to the reasons of such a distribution, but it is possible that I1's interpretations are influenced both by the source language and the style of Spanish and French original speakers, which might be different from the German ones.

Table 48. The proportion of high frequency words analysed on 720 word corpora of interpreter I1's non-interpreted and interpreted discourse consisting of text samples.

	I1_ORG_SP_EN	I1_SI_FR_EN	I1_SI_ES_EN	I1_SI_DE_EN
% High frequency words	63%	55%	59%	62%

Table 49. Lexical sophistication of I1's interpreted and non-interpreted discourse: Results of Chi-square tests calculated on raw counts in 2x2 contingency tables.

	I1_ORG_SP_EN vs. I1_SI_FR_EN	I1_ORG_SP_EN vs. I1_SI_ES_EN	I1_ORG_SP_EN vs. I1_SI_DE_EN
χ^2	8.99	2.28	0.19
p-value	$p<0.01$	$p>0.05$	$p>0.05$

The comparison of the range of vocabulary used by I2 in interpretations and non-interpreted discourse proves that the texts are very homogenous (see Table 50 and Table 51). Not only there is no difference between the interpretations from French and German and non-interpreted discourse, but also between the interpretations from the two respective languages ($\chi^2=0.28$, $p>0.05$). This suggests, that with respect to the range of vocabulary

I2's speaking style does not differ from his interpreting style and his use of vocabulary is not easily affected by interference.

Table 50. The proportion of high frequency words analysed on 720 word corpora of interpreter I2's non-interpreted and interpreted discourse consisting of text samples.

	I2_ORG_SP_EN	I2_SI_FR_EN	I2_SI_DE_EN
% High frequency words	62%	57%	58%

Table 51. Lexical sophistication of I1's interpreted and non-interpreted discourse: Results of Chi-square tests calculated on raw counts in 2x2 contingency tables.

	I2_ORG_SP_EN vs. I2_SI_FR_EN	I2_ORG_SP_EN vs. I2_SI_DE_EN
χ^2	3.73	1.96
p-value	p>0.05	p>0.05

Interpreter I1 and I2 do not differ significantly with respect to the incidence of high frequency words neither in interpretations from French ($\chi^2=1.77$, p>0.05) nor from German ($\chi^2=3.13$, p>0.05). Neither is the overall tendency contradictory: in general in both cases interpretations are slightly more lexically sophisticated i.e. less simplified. What both interpreters differ in is the homogeneity of their texts: I1's interpretations from French deviate significantly from his non-interpreted discourse, while I2's interpretations and non-interpreted texts do not differ significantly.

4.3. Explicitness

The examination of explicitness in the present paper has so far involved the analysis of frequency of optional *that* connective after reporting verbs, the frequency of linking adverbials and the frequency of apposition markers. Frequency of both linking adverbials and apposition markers turned to be exceedingly small already in oral subcorpora of TIC and was a major roadblock to the full verification of explicitness hypothesis. Hence, due to insufficient data the tendency to explicitness among individual interpreters can only be analysed based on the frequency of optional *that* connective after reporting verbs. It must be stressed, however, that also this analysis has been carried out on a very small and unbalanced sample of interpreters' output, and therefore it is impossible to venture any generalisations. For the same reason, no statistical test has been applied in this section.

4.3.1. Optional *that*

In the analysis of optional *that* it is important to examine not only the frequency of its factual occurrence, but also the number of times that it could have been used, but was not. Hence Table 52 and Table 53 show not only the normalised results, but also the raw frequencies and percentage values expressing the proportion of the number of occurrences of *that* connective to the actual number of situations, in which it could have been used. It must be stressed, however, that the analysed samples are very small and such are the observed frequencies of the reporting verbs. The percentage values presented in Table 52 and Table 53 are calculated based on raw values and should serve as indicators only, as at such low frequencies the percentage values change drastically with any additional occurrence.

The use of optional *that* in interpreter I1's spoken discourse is considerably less frequent when compared to his interpretations. The samples are very small, so it is impossible to speak of any definite trends, but the lower proportion of optional *that* in I1's spoken discourse signifies small tendency to explicitness, which in this case may be probably augmented by a slightly lower level of formality. As it was suggested in Section 3.1.2.1. the higher level of formality might be one of the key factors contributing to the higher frequency of optional *that*. The researcher's subjective impression was that I1's speaking style was slightly less formal than his interpretations and than I2's speaking style. Was it in fact the case, such an outcome could lend support to the claim that the level of formality may be an important factor contributing to the frequency level of this optional complementizer.

Table 52. The proportion of optional *that* after reporting verbs analysed on the full collection of interpreter I1's interpretations and the full transcript of non-interpreted discourse normalised per 1000 words. Raw frequencies in brackets.

	I1_ORG_SP_EN		I1_SI_FR_EN		I1_SI_ES_EN		I1_SI_DE_EN	
	that	z.c.	that	z.c.	that	z.c.	that	z.c.
Total	0.37(2)	2.02 (11)	2.64 (5)	1.59 (3)	1.35 (4)	1.69 (5)	1.42 (7)	0.61 (3)
that %	15%		63%		44%		70%	

I1's spoken discourse seems to be altogether less explicit than any of his interpretations, which, on the other hand, do not form a homogenous group. Admittedly, they are also not homogenous in the proportion of source speeches delivered in oral and read mode: in case of interpretations from French about 28 % of the source speeches were

spoken, from Spanish 50% and as from German much as 70%. One could intuitively assume that the corpus, in which the vast majority of the source speakers spoke in an unscripted fashion would most resemble the corpus of I1's non-interpreted discourse. By analogy, it would seem natural that the interpretations from French should be at the one end of the continuum and interpretations from German on the other. As the frequency of the optional complementizer lends support to neither of the above assumptions, the impact of the mode of delivery cannot be deemed a decisive factor for this parameter.

The proportion of optional *that* in interpretations from French and German is relatively similar and comparatively higher than in interpretations from Spanish. This is a completely different pattern from the one observed in TIC, where interpretations from Spanish are characteristic of the highest proportion of optional complementizer *that* of all interpretations. That interpretations from Spanish stand out in this set in a way that is completely different from the tendencies observed in TIC may be caused by manifold factors. As suggested earlier in this Chapter, it may be that interpreter I1 imitates the original speakers' style, which happens to be less explicit/formal. Another viable explanation involves risk management strategies. In his account of the differences between interpretations from different languages (see Section 4.1.) interpreter I1 argued, that for him, it was not the structure of the language, but his own level of confidence in his language skills that counted most. More precisely, he admitted, that although he has spoken German and French only a few years longer he felt more confident while interpreting from these two languages. This situation can be associated with Pym's (2007: 11-12) idea of risk-reducing strategies (see 1.5.2.). From his assumptions, it follows that when less confident interpreters would rather try to refrain from being specific and explicit. This particular case, does not involve any real incompetence, as both I1 and I2 are very experienced in interpreting in all their language combinations and they both have been accredited by the EU institutions that are renown for high interpreting standards. Therefore, the slightly lower level of confidence in interpretation from one of the source languages would probably be rather a subjective opinion than an actually lower standard of interpretation. This subjective assessment may, however, trigger more risk aversion strategies contributing to the lower level of explicitness.

Interpreter I2 more often resorts to optional *that* than to zero connective, both in non-interpreted discourse and interpretations. Had the mode of delivery of the source speeches had a significant impact on the frequency of the optional complementizer in the target text, then similar results should be expected across I2's both interpretation corpora

(as both of them have a very similar proportion of the speeches that were originally spoken or read out: 1/3 to 2/3. See Figure 13). This, however, does not occur. Instead, the frequencies are completely different. The decisive factor must be also different from the source language, as the corpus of I1's interpretations from the same language i.e. French consists of the same proportion of interpretations of speeches that were originally read out or spoken (1/3 to 2/3. See Figure 12) and yet the frequency of optional *that* seems to be different.

As the mode of delivery does not appear to exert a great impact on this parameter, it is necessary to discuss the effect of other factors. In interpretations from French I2 chose to use the optional connective in all cases. Considering his comment about transcoding (see Section 4.1.) and the fact that a complementizer is in French obligatory after reporting verbs it may be assumed that such a result could be caused by interference.

Table 53. The proportion of optional *that* after reporting verbs analysed on the full collection of interpreter I2's interpretations and the full transcript of non-interpreted discourse normalised per 1000 words. Raw frequencies in brackets.

	I2_ORG_SP_EN		I2_SI_FR_EN		I2_SI_DE_EN	
	that	z.c.	that	z.c.	that	z.c.
TOTAL	1.21 (7)	0.35 (2)	1.96 (4)	0.00 (0)	3.16 (14)	1.81 (8)
that %	78%		100%		64%	

Interestingly, however, with respect to the analysed parameter I2's non-interpreted discourse classifies somewhere in between interpretations from French and German. Such a result also contradicts the patterns recurring in oral subcorpora of TIC. Although the differences are not enough to prove their statistical significance, to corroborate the hypothesis that interpretations are always more explicit than non-interpreted discourse one would expect an opposite trend.

The examined corpora of I1 and I2's interpreted and non-interpreted discourse are very small and the results of such an analysis may only be tentative.

Compared to Interpreter I2, Interpreter I1 uses the optional complementizer far less frequently while speaking and on average also less frequently in interpretation. This suggests that he may be in general less explicit than interpreter I2 i.e. that his speaking style would also affect his interpretation with this respect. Additionally, the interview with both interpreters provides viable answers that could explain the differences between them. Interpreter I1 seems to focus more on adjusting his speaking style to the original speaker, while Interpreter I2 is more formal while interpreting and admits that it is possible to

occasionally transcode the message from French word-for-word. They may of course both be using the same strategies, but in different situations and with varying frequency. Their focus on the 'favoured' strategy could partially be responsible for the observed results: I1's interpretations are more varied, while I2's more formal with interpretations from French characteristic of higher frequency of optional *that*. Based on the results of the above analysis it is visible, however, that none of the interpreters is particularly consistent with the use of optional *that* across interpretations from different languages and that either due to interference, level of confidence in their interpretation, level of formality or yet other factors, the incidence of optional connective seems to vary. This effect may also be ascribed to the size of the corpora.

4.4. Normalisation

Normalisation in the present study has been so far investigated with two measures: the frequency of common lexical bundles and the frequency of fixed expressions most characteristic of spoken professional English. Due to a limited size of samples in the analysis of interpreting style normalisation can be examined based on only one of the two parameters that have been adopted earlier in this study i.e. the most frequent fixed expressions in spoken professional English.

4.4.1. Fixed expressions

The investigated fixed expressions are deemed to be characteristic of professional spoken English and do not pertain to any particular topic. The list does not contain any pure idioms and comprises mostly prepositional phrases: *as far as*, *as to*, *as well*, *as well as*, *at all*, *first of all*, *in fact*, *in favour of*, *in order to*, *in terms of*, *kind of*, *of course*, *on behalf of*, *once again*, *sort of*, *with regard to*. It is assumed that a higher frequency of these fixed expressions in a corpus points to the use of more conventional language i.e. the tendency to normalisation.

Interpreter I1 seems to use fixed expressions more frequently in his interpretations than in non-interpreted discourse. The difference is not large between interpretations from

French and German, but the frequency of investigated items is more than twice higher in interpretations from Spanish than in I1's non-interpreted speech. Such pattern is in line with the normalisation hypothesis, which stipulates that a more conventional language is used in interpretation. This hypothesis has, however not been confirmed in the analysis carried out on TIC and due to a small size of the investigated samples it is impossible to predict, if the trend could also hold for other interpretations delivered by I1.

Table 54. The frequency of fixed expressions in I1's interpreted and non-interpreted discourse normalised per 1000 words.

	I1_ORG_SP_EN	I1_SI_FR_EN	I1_SI_ES_EN	I1_SI_DE_EN
TOTAL	3.12	4.23	7.08	4.26

The trend observed in the output of Interpreter I2 is completely different. In his case the most conventional language seems to be used in non-interpreted discourse. I2's interpretations from French contain slightly lower number of frequent fixed expressions than spoken discourse and twice as many as interpretations from German. The data of Interpreter I2 does not corroborate normalisation hypothesis and also at the same time deviates from the pattern observed in TIC. In Translation and Interpreting Corpus the frequency of examined fixed expressions was comparable in the subcorpus of native English speeches and the subcorpora of interpretations from Germanic languages. In this case, non-interpreted discourse is most conventional and interpretations are not homogenous.

Table 55. The frequency of fixed expressions in I2's interpreted and non-interpreted discourse normalised per 1000 words.

	I2_ORG_SP_EN	I2_SI_FR_EN	I2_SI_DE_EN
TOTAL	5.53	4.40	2.26

Both interpreters appear to use the examined expressions differently. While a slightly higher frequency was observed in I1's interpretations rather than in his non-interpreted discourse, the trend observed in I2's corpora is completely opposite. Interestingly both interpreters reach for frequent expressions with a similar frequency when interpreting from French, but this is where the similarity ends. While interpreting from German I1 uses the investigated fixed expressions almost twice more than I2. On the other hand, I2 uses them more often in non-interpreted discourse.

4.5. Simplification, explicitness and normalisation as idiosyncratic stylistic features

It appears that certain parameters investigated in search of tendencies to simplification, explicitness and normalisation may indeed aid in revealing idiosyncratic stylistic preferences.

As apparent from the analysis presented in the previous section interpreter I1 does not seem to manifest any consistent tendency to lexical simplification or complexity that could be discernible in all his interpretations from different languages. Apart from subtle differences, usually not big enough to earn statistical significance (with one exception, see Section 4.2.1.3.) I1's interpretations and non-interpreted discourse appear to be very similar. In fact no statistically significant tendency to simplification was observed in any of his interpretations. On the contrary, I1's interpretations from French do stand out from other corpora as more lexically sophisticated due to a lower proportion of high frequency words (see 4.2.1.3.). As regards tendencies to simplification, I1's interpretations are rather consistent with his speaking style, in that they do not deviate significantly from the non-interpreted discourse (except for the abovementioned tendency to lexical sophistication in interpretations from French). However, there seems to be a statistically significant difference with respect to repetitiveness and lexical sophistication between I1's interpretations from French and German, which suggests a significant impact of interference and lack of homogeneity with this respect in his interpretations.

The analysis of the tendency to lexical simplification proves that interpreter I2 is fairly consistent with language used in interpretations and non-interpreted discourse and that his interpretations are never lexically simplified. The only parameter of simplification that clearly points to a difference between I2's non-interpreted and interpreted discourse is the analysis of list heads, which indicates that regardless of the source language I2 is significantly less repetitive while interpreting. The analysis of parameters pointing to simplification has not shed light on any significant differences other than greater lexical repetitiveness in non-interpreted discourse.

It has been ascertained that interpreters I1 and I2 are different notably in one aspect: while I1 does not exhibit any consistent patterns of repetitiveness, informativeness or lexical sophistication that could distinguish all his interpretations from non-interpreted discourse, I2 does prove to be less repetitive in all his interpretations regardless of the

source language. As evident from the analysis in the previous sections, one of the parameters helped to spot one of the characteristic traits of I2's interpreting style.

None of the two interpreters is particularly consistent in the tendency to explicitness. I1 uses optional *that* on average more frequently in his interpretations than in non-interpreted discourse, yet his interpretations from Spanish seem to be considerably less explicit than interpretations from French and German. I2 uses reporting verbs with *that* more frequently than with zero connective, but demonstrates no consistency with this respect and no greater explicitness in interpreting than in non-interpreted speeches. When compared to I2, I1 seems to be slightly more consistent in the overall tendency to explicitness in interpreting as opposed to non-interpreted discourse. I2, on the other hand, seems to be more explicit both in interpretation and non-interpreted discourse.

Undoubtedly the size of examined samples hinders the interpretation of data. At such small frequencies it is difficult to speak of any consistency. The tendency to normalisation of Interpreter I1 seems to be stronger in his interpretations, but the observed results suggest that it is dependent on the source language. A completely opposite trend characterises Interpreter I2, who uses more conventional language while speaking. What both I1 and I2 have in common is that their interpretations from different source languages vary. The non-interpreted discourse of both interpreters seems to stand out from their interpretations either as the least normalised (I1) or the most normalised (I2), in both cases on the end of the normalisation continuum rather than between the values recorded for respective interpretations. This suggests that speaking style of both interpreters differs from their interpreting style.

4.6. Style vs. universal tendencies

As regards simplification, the analysis of idiosyncratic differences undermines the observations regarding the universal patterns in interpreting in all examined aspects: repetitiveness, informativeness and range of vocabulary. Firstly, while in TIC, texts interpreted from Spanish, German and Dutch are as repetitive as native English speeches and less repetitive than the ones interpreted from French, interpreter I2 proves to be significantly less repetitive in interpretations from French and from German than in non-interpreted discourse. In this way I2 manifests a feature of his individual interpreting style

that is significantly different from global patterns observed in TIC. Secondly, in TIC, native English speeches demonstrate lower lexical density than all examined interpreting subcorpora, whereas both I1 and I2 seem to use the language characteristic of the same level of informativeness in all examined texts. Finally, although in TIC the largest proportions of high frequency words can be found in native English speeches, pointing to their lexical simplification, I2 uses high frequency words equally often in all interpretations as well as in non-interpreted discourse.

The investigation of the tendency to explicitness shows that the results observed in the output of individual interpreters deviate from the global patterns. With his non-interpreted discourse being far less explicit than the interpretations, Interpreter I1 remains in line with the general tendency observed in oral subcorpora of TIC (however the discrepancy between non-interpretations and interpretations is considerably more pronounced in this case). The pattern is distorted in that I1's interpretations from Spanish are the least explicit and there is a difference between interpretations from French and German, while in TIC interpretations from Spanish are the most explicit and interpretations from German and French do not differ in this respect. The outcomes of the analysis of Interpreter I2's speaking and interpreting style do not correspond with the global patterns observed in TIC either. The frequency of the optional complementizer *that* in I2's speaking style happens to be higher than in his interpretations from German and lower than in the texts interpreted from French, which translates into his being more explicit while speaking than while interpreting from German. Thus, the tendency of interpretations to be more explicit than native target language speeches cannot be deemed universal. There remains of course a pertinent question, if investigation of larger samples of similar texts by the same authors would confirm the pattern. It is especially valid since the results obtained in Section 4.3.1. due to their low numerical value cannot be tested statistically.

The outcome of the analysis investigating the tendency to normalisation in interpreting does not reflect the patterns observed in TIC. Interpreter I1 appears to use more normalised language in his interpretations, while Interpreter I2 seems more inclined to do so in his non-interpreted speech. The texts produced by neither reflect exactly the patterns visible in TIC, where a higher frequency of examined fixed expressions pointing to greater normalisation is observed in interpretations from Romance languages, while interpretations from Germanic languages do not deviate significantly from native English speeches with this respect. The analysis of I1's interpreted and non-interpreted texts seem to corroborate

normalisation hypothesis, while the examination of I2's data contradict it. None of those results, however, can ultimately prove or disprove the hypothesis due to insufficiently big samples.

4.7. Limitations of the micro study of idiosyncratic variation: interpreting style

The study described in the present Chapter provides valuable insight on tendencies that can be observed in interpretations from different source languages and non-interpreted speech of individual interpreters. It must be stressed, however, that the analysis has been carried out on a compilation of data that is not in every aspect ideal.

The first limitation regards the size of samples subject to analysis. The number of tokens of all texts produced by individual interpreters exceeds ten thousand words, which is in itself a lot for a study on interpretation considering the fact that transcriptions of authentic interpreting data is very difficult to obtain. On the other hand, corpus-based methodology requires abundance of text for a hypothesis to be properly tested and the results statistically significant. This criterion has not been met and therefore, the results presented in Chapter Three should be treated as indicative only.

The second constraint is very closely related to the size of the examined corpora, authenticity of data and difficulty in obtaining relevant interpreting material and regards balance. It is very difficult to maintain balance of a variety of factors such as source language speaker, topic or mode of delivery within such small corpora. In fact, in the investigation of style presented in this Chapter, only texts used for the analysis of simplification were properly balanced with respect to the source speaker and mode of delivery of the original speech. In the remaining analyses of explicitness and normalisation all available texts were used as the list of investigated lexical items was far more specific.

The third flaw of the data is that non-interpreted recordings were carried out for the sake of the experiment and the situation of the recording was less formal. What is more, the interpreters were prompted by questions to encourage them to talk on the subjects related to the European Union. Although the subjects were not interrupted in their responses, which therefore assumed the form of at least two-minute monologues, the mere fact that they were answering a question could have affected the structure of the text. It is believed, however, that such a constraint would not have severely affected lexis. What it could have affected

was the level of formality and possibly the tendency to repetition. These aspects have been taken into account in the interpretation of the results.

Finally, as in any study in the field of translation and Interpreting Studies conducted on comparable corpora it is impossible to verify to what extent certain results could have been influenced by the source texts.

For the reasons enumerated above, the results obtained in the study presented in Chapter Four must be treated as indicative only. In order to confirm the observed trends, the analysis should be carried out on a far bigger and more representative sample.

Table 56. Juxtaposition of the outcomes of micro analysis.

Examined universals	Tested parameters	Are individual interpreters consistent in their tendencies regardless of the source language?	Is individual interpreting style different from speaking style?		Do individual interpreters differ from each other in their interpreting style?
			I1	I2	
SIMPLIFICATION	List head	yes	no	yes (interpreted texts are less simplified)	yes
	Lexical density	no	no	no	no
	High frequency words	no	yes* (interpreted texts are less simplified)	yes* (interpreted texts are less simplified)	no
EXPLICITNESS	Optional connective <i>that</i>	yes	yes (interpreted texts are more explicit)	no	yes
	Linking adverbials	-	-	-	-
	Apposition markers	-	-	-	-
NORMALISATION	Lexical bundles	-	-	-	-
	Fixed phrases	yes	yes (interpreted texts are more normalised)	yes (interpreted texts are less normalised)	yes

* denotes an observable, but not statistically significant trend; “-“ signifies that the parameter has not been tested due to small size of the sample.

Conclusions

As it has been argued in Chapter One, the research on universals and investigations of style complement one another. Universality of a selected feature may be questioned if it cannot be traced in texts produced by an individual and it is difficult to discern a feature that could characterise the style of only one particular individual without comparing it to a broader context of general trends. Although this relation has so far not been extensively researched in the field of Corpus-based Translation or Interpreting Studies, such is the perspective assumed in the present dissertation, whose objective is to investigate whether simplification, explicitness and normalisation, commonly acknowledged translation universals may be regarded as interpreting universals and to what extent their prominence in interpreted discourse depends on interpreter's idiosyncratic stylistic preferences.

This idea has been operationalised into four research questions and four preliminary hypotheses that have been investigated in macro and micro analyses introduced in Chapter Two and carried out in Chapter Three and Chapter Four. The present concluding section reviews the outcomes of the macro analysis and the micro analysis in a consolidated way and against this backdrop strives to provide answers to the research questions. To facilitate the analysis, the results of tested parameters are juxtaposed in Table 57.

Table 57. Juxtaposition of the outcomes of macro and micro analysis.

Examined features	Tested parameters	MACRO ANALYSIS			MICRO ANALYSIS			
		Has the universal been confirmed?			Are individual interpreters consistent in their tendencies regardless of the source language?	Is individual interpreting style different from speaking style?		Do individual interpreters differ from each other in their interpreting style?
		TIC	I1	I2		I1	I2	
SIMPLIFICATION	List head	no	no	no (all interpreted texts are less simplified)	yes	no	yes (interpreted texts are less simplified)	yes
	Lexical density	no (all interpreted texts are less simplified)	no	no	no	no	no	no
	High frequency words	no (all interpreted texts are less simplified*)	no (all interpreted texts are less simplified*)	no (all interpreted texts are less simplified*)	no	yes* (interpreted texts are less simplified)	yes* (interpreted texts are less simplified)	no
EXPLICITNESS	Optional connective <i>that</i>	yes (interpreted texts are more explicit)	yes (interpreted texts are more explicit)	no	yes	yes (interpreted texts are more explicit)	no	yes
	Linking adverbials	no	-	-	-	-	-	-
	Apposition markers	yes* (interpreted texts are more explicit)	-	-	-	-	-	-
NORMALISATION	Lexical bundles	yes (interpreted texts normalised)	-	-	-	-	-	-
	Fixed phrases	no	yes (interpreted texts are more normalised)	no (interpreted texts are less normalised)	yes	yes (interpreted texts are more normalised)	yes (interpreted texts are less normalised)	yes

* denotes an observable, but not statistically significant trend; “-“ signifies that the parameter has not been tested due to small size of the sample.

None of the hypothesized universals has been uniformly confirmed by all tested parameters and therefore the first hypothesis stipulating that all simultaneous interpretations tend to use more simplified, more explicit and more normalised language has been disproved.

The first of the investigated parameters of simplification i.e. the proportion of the corpus represented by list head pointing to the level of repetitiveness does not unequivocally point to simplification as a universal feature of interpreting. The outcomes of macro analysis suggest that the tendency to repetitiveness in interpreting may be influenced by several factors such as the mode of delivery of the original speaker or the source language. The analysis of TIC shows that interpretations from French seem to be the least repetitive – even less repetitive than the native English speeches, while interpretations from Spanish and German the most. A similar trend can also be observed in the micro analysis of I1's interpretations. On the other hand the macro analysis of interpretations from Spanish focused on the impact of the mode of delivery of the original speaker proves that at least in the case of this language combination the mode of delivery may have a significant impact on the repetitiveness of interpreted language: the interpretations of speeches that were originally read out prove to be as repetitive as native English speeches, while interpretations of unscripted speeches are significantly more repetitive. It is difficult, however, to predict if that trend could be universal, as interpretations from Spanish in all analysed corpora seem to be always the most repetitive of all interpretations. Micro analysis also does not confirm the simplification hypothesis. Interpretations delivered by I1 are not statistically different from his non-interpreted discourse. The comparison of percentage values that could suggest certain trends shows that I1's interpretations from different languages are either more repetitive (from Spanish and German) or less repetitive (from French) than his non-interpreted discourse thus even if the percentage values were indicative, they would suggest that the interpretations differ with respect to the source language and are not as a group either more or less repetitive than I1's non-interpreted discourse. The case of I2 is completely different. His interpretations are very homogenous with this respect and statistically significantly less repetitive than his non-interpreted discourse. This trend might be influenced by the mode of delivery of the original speeches, which in case of the corpora used in the test for simplification was always read rather than spoken. In terms of a universal pattern, such an outcome points rather to a trend opposite to simplification. As a result, judging by macro and micro analyses described in Chapter

Three and Chapter Four, repetitiveness and cannot be considered a universal feature of interpreting. It has been confirmed that when interpretations are based on speeches that were originally read out, interpreted language is likely to be less repetitive than non-interpreted language.

The examination of lexical density does not corroborate the simplification hypothesis either. Macro analysis points to a uniform and statistically significant trend that holds for all interpreting corpora in TIC according to which interpretations are more lexically dense i.e. more informative than speeches originally produced in English. Moreover, the investigation of interpretations of Spanish speeches delivered in two different modes proves that the mode of delivery of the original speech does not have a statistically significant impact in this language pair. Based on the outcome of macro analysis one could infer that regardless of the source language and the mode of delivery of the original speech, interpretations are always significantly more dense than native English speeches. As stressed in Chapter Two, a feature could only be regarded universal if proved to be a feature of all investigated corpora and the micro analysis of texts produced by interpreters I1 and I2 presented in Chapter Four does not confirm this trend. The language of non-interpreted discourse of interpreter I1 does not differ statistically from his interpretations. Judging merely by percentage values, his interpretations from French and Spanish are less lexically dense than interpretations from German and non-interpreted discourse, which if treated as indicative of a trend would point to the lack of homogeneity in the group of I1's interpretations. Exactly the same pattern can be observed in case of interpreter I2: the statistical test does not point to any differences between interpreted and non-interpreted texts and percentage values could only suggest lack of homogeneity in the group of interpretations. Hence, although the macro analysis points clearly to a uniform pattern, according to which interpretations are always more lexically dense i.e. more informative and thus less simplified than native English speeches, the feature cannot be regarded universal, as it has not been reflected in micro analysis.

The final parameter of simplification involves the examination of the proportion of high frequency words which indicates the level of lexical sophistication. The higher the proportion of the investigated lexis, the more simplified the text. Of all parameters of the putative universals investigated in this study, greater lexical sophistication is most likely to be a universal feature of interpreting, although the tendency has not been fully corroborated by all statistical tests. The results obtained in macro analysis of oral subcorpora of TIC

prove unambiguously that all interpreting subcorpora contain a lower proportion of high frequency words than native English speeches. This clearly disproves the simplification hypothesis and suggests that interpretations are more lexically sophisticated regardless of the source language. The examination of interpretations of Spanish speeches read out vs. spoken prove that the tendency may be stronger than the impact of the mode of delivery of the original speeches. The same trends can be observed in micro analysis of interpreted and non-interpreted discourse of individual interpreters I1 and I2. The percentage values indicate that interpretations are more lexically sophisticated than non-interpreted discourse. The degree of this sophistication does vary, but in the case of both I1 and I2 the corpora of non-interpreted discourse are characterised by a higher proportion of high-frequency words than any of the interpretations. The trend, however, is not corroborated by the test of statistical significance.

The present study investigated the tendency to explicitness by means of three parameters: the frequency of optional complementizer *that*, the incidence of linking adverbials and apposition markers. Only the first parameter could be investigated both in macro and micro analysis. The macro analysis reveals that optional complementizer *that* follows reporting verbs more frequently in all interpreting subcorpora of TIC, regardless of the source language proving greater explicitness of the interpreted language (which may be caused by a greater formality see Section 3.1.2.1.). The micro analysis does not corroborate these results. Indeed, in the case of interpreter I1 interpreted language seems to be more explicit; the non-interpreted language shows a clear preference for the zero connective, while interpretations from French, Spanish and German exhibit a higher incidence of optional complementizer *that*. The tendency, however, is manifested in a very uneven way with the interpretations from Spanish manifesting a lower frequency of the investigated phenomenon when compared to the interpretations from French and German. The analysis of the output of interpreter I2 contradicts the explicitness hypothesis. While interpreting from French I2 opted for the optional *that* in all cases, where the connective could have been used, in the case of interpretation from German *that* was used in only 63 per cent of the total instances that allowed for it. Finally, in non-interpreted discourse I2 resorted to optional *that* in 78 per cent of instances, which disrupts the pattern observed in TIC and in I1. Along with the prerequisites set forth in Chapter Two, for a tendency to be regarded universal, it has to be corroborated by all of the tests carried out in the present study. Hence, it cannot be ultimately confirmed that the language of simultaneous

interpretation would as a rule manifest a higher incidence of optional connective *that* after reporting verbs.

The other two parameters that indicate a greater tendency to explicitness i.e. the frequency of linking adverbials and the incidence of apposition markers could only be investigated in macro analysis due to their lower overall frequency in the spoken language, which would in the case of micro analysis be insufficient. The examination of the frequency of linking adverbials in macro analysis shows that the outcomes of this parameter rely heavily on the source language. It has been proved that when compared with the subcorpus of native English speeches, the investigated linking adverbials occur statistically significantly more frequently in the interpretations from French and from German, while the interpretations from Spanish and from Dutch do not differ from them significantly. The frequency of apposition markers in interpretations from French, Spanish and Dutch is statistically significantly different from the one observed in native English speeches. There is much smaller gap between the latter and interpretations from German, but it must be stressed that the difference is very close to being statistically significant. Although the incidence of apposition markers exhibited in interpreting subcorpora of TIC cannot indisputably point to greater explicitness of the interpreted language, it is likely that a bigger sample could confirm such a tendency. From the analysis of the compiled corpora it follows that the tendency is not pronounced enough to be regarded universal.

The tendency to normalisation has in the present study been investigated with two parameters: the frequency of common lexical bundles and fixed expressions most characteristic of professional spoken English. The examination of common lexical bundles in macro analysis confirmed the normalisation hypothesis. Lexical bundles generated based on the subcorpus of native English speeches and common for all oral subcorpora of TIC prove to be significantly more frequent in interpreting subcorpora than in the corpus based on which they have been generated. Although the tendency is statistically significant across all interpreting corpora, the difference between native English speeches and interpretations from French is least pronounced and, in fact, the latter resembles the reference subcorpus more than any of the remaining interpreting subcorpora. As the same trend can be observed in the corresponding examination of translation subcorpora, it may be assumed that interference greatly affects this parameter, but its impact is still less significant than the tendency to use conventionalised language. Lexical bundles were not investigated in the micro analysis of interpreting style due to small size of samples. It would be, however

interesting to know if the same pattern could be observed in individual interpreters. Were that the case, this parameter of normalisation would be a likely candidate for a universal feature.

The second parameter does not corroborate the normalisation hypothesis. The investigated set of fixed expressions characteristic of spoken professional English turns to be best represented in interpretations from Romance languages, while interpretations from Germanic languages exhibit the same incidence of the investigated expressions as native English speeches. Interestingly, the pattern in translated subcorpora of TIC is almost directly reverse: the occurrence of investigated fixed expressions is higher in translations from Germanic languages. For that reason, it is difficult to ascribe the lack of homogeneity in the interpreting and translation subcorpora, which originate from the same source text merely to interference. One of the possible reasons may be the interpreters' stylistic preference.

The second preliminary hypothesis assumed that individual interpreters are consistent in all their tendencies to simplify, explicitise and normalise regardless of their source language. It has been only partially confirmed: while the interpreters are consistent in some tendencies (e.g. to be less repetitive while interpreting from different languages than while speaking) and inconsistent in others.

There are three features that appear to distinguish I1's interpretations in general from his non-interpreted discourse; greater lexical sophistication (no tendency to simplification), greater explicitness and more normalised language. None of these tendencies have been confirmed by statistical tests and they do not manifest at a consistent level across all interpretations. Of all texts produced by I1, non-interpreted discourse exhibits the greatest incidence of high frequency words, indicating that the remaining text i.e. all interpretations are more lexically sophisticated. The group of interpretations is, however, very heterogeneous with this respect (see Section 4.2.1.3.), with some interpreting subcorpora (SI_ES_EN) resembling non-interpreted discourse more than other interpretations. Furthermore, all texts produced by I1, be it interpretations or non-interpreted discourse, are fairly consistent with respect to repetitiveness and informativeness, which proves that the source language does not influence these two features and that I1's interpretations are not simplified. All I1's interpretations seem also to use optional complementizer *that* on average more often than non-interpreted texts. It does not follow, however, that their level of explicitness is consistent: I1's interpretations from

Spanish stand out as less explicit. Finally, I1 appears to use more normalised (conventional) language while interpreting: his interpretations exhibit greater incidence of the investigated fixed expressions than non-interpreted texts. Again this feature is far from consistent: while interpretations from French and German are very similar with this respect, interpretations from Spanish manifest almost twice as high frequency of the investigated expressions suggesting that the language is most normalised.

A tendency to lower repetitiveness, greater lexical sophistication and less normalised language make I2's interpretations different from his non-interpreted speech. While interpreting I2 has proved to be less repetitive and seems to use more sophisticated vocabulary and less normalised language. Statistical tests point to a significant difference in repetitiveness of language used by I2 in his interpretations from French and German when compared to non-interpreted discourse proving that his interpretations are less simplified. Apart from using less repetitive language, I2 seems also to employ a smaller number of high frequency English words in his interpretations than in non-interpreted discourse. This trend is not strong enough to be confirmed by statistical tests i.e. neither interpretations from French, nor from German differ significantly from the non-interpreted discourse, but the percentage values representing the proportion of high frequency words in these corpora are consistent and lower than in the non-interpreted speech. If indeed such a tendency could be proved on a bigger sample as a consistent trait of I2's interpreting style, it would be less pronounced than I2's tendency to lower repetitiveness and more difficult to spot. The last parameter of simplification i.e. lexical density shows that regardless of the source language all interpretations manifest a consistent level of informativeness, which, however, does not deviate from the one characteristic of I2's non-interpreted discourse.

Judging by normalised frequencies of the investigated expressions characteristic of professional spoken English it may be inferred that I2 employs more such expressions while speaking than in interpretations. Nevertheless, with the occurrence of the abovementioned expressions twice higher in interpretations from French than in interpretations from German one cannot speak of consistency that could be independent from the source language.

The third preliminary hypothesis has been confirmed: the two interpreters, whose interpreting and speaking style has been subject to analysis do differ in the way they interpret and the distinctive features involve the tendency to repetitiveness, explicitness and normalisation.

Firstly, neither I1 nor I2 shows a consistent tendency to simplify language when interpreting. I2 is consistently less repetitive in simultaneous interpretation than in non-interpreted discourse. This feature is manifested in his interpretations both from French and from German and both interpreting corpora are highly homogenous with this respect. At the same time each deviates significantly from the corpus of I2's non-interpreted discourse. No such consistency could be spotted in texts interpreted by I1, none of which deviates from his non-interpreted speech. At the same time, both interpreters differ significantly in their interpretations from German and their level of repetitiveness is very similar in interpretations from French. This means that when interpreting from French, I1 is as repetitive as I2, but that is not a consistent feature of his interpreting.

Secondly, I1 seems to be more explicit in his interpretations than in non-interpreted speech, while I2 is very inconsistent with this respect. Due to a small size of the tested sample and consequently, a low number of occurrences, it is impossible to determine how strong this trend is. The respective percentage values show, however, that I2 would rather opt for a zero connective after reporting verbs in non-interpreted discourse and choose optional *that*, at least twice as frequently in all interpretations. I2, on the other hand, prefers optional *that* both in interpretations and in non-interpreted discourse, but the pattern of use in interpretations does not deviate from non-interpreted texts in any consistent way.

Finally, the two interpreters differ in their tendency to normalisation. While I1's interpretations tend to use more conventional language i.e. more high frequency fixed expressions, interpretations delivered by I2 exhibit a lower incidence of the said items than his non-interpreted discourse. Similarly to explicitness, this difference has not been verified by statistical tests due to low frequency of investigated items, but comparison of normalised frequencies point to a trend in I1's interpretations, which cannot be spotted in I2's interpreted discourse.

Of the three features that distinguish I1 and I2's interpreting styles, one has been confirmed as statistically significantly different and the remaining two should be tested on bigger samples⁸. Be it just one, two or all three discussed features that make it possible to tell the two interpreters apart, the fact that such a distinction can be made proves that interpreters have their own interpreting style, which may, in reality, be manifested by more characteristic features than the ones investigated in this study.

⁸ The tests have not been carried out at all due to low frequency of the investigated items.

The final, fourth, preliminary hypothesis stipulating that one's interpreting style is similar to their speaking style has been partially confirmed. Looking at the tests for statistical significance carried out to measure differences between corpora compiled of I1's interpretations from three different languages and his non-interpreted speech it is impossible to spot any pronounced tendency consistent in all interpretations that could significantly differ from non-interpreted speech. This means that among the investigated features none is prominent enough to stand out as characteristic of I1's interpreting or speaking style. On the other hand, percentage values calculated for all respective parameters, may be indicative of certain less pronounced tendencies. These would involve a greater use of optional complementizer *that* in interpretations in general (although interpretations do not seem to be homogenous with this respect) and higher incidence of fixed expressions characteristic of spoken professional English. The latter two tendencies, would suggest slight tendencies to explicitness and normalisation.

It appears that with respect to the proportion of high frequency words and the use of optional *that*, it is interpretation from Spanish that resembles I1's non-interpreted discourse the most. On the other hand, his interpretations from Spanish manifest the highest incidence of fixed expressions symptomatic of normalisation. This may be associated with a comment made by the interpreter himself, in which he admitted to feeling less confident in his interpretations from Spanish. One could only speculate that being less convinced of his interpretation I1 may attempt, consciously or subconsciously, to sound most naturally, which would explain the proportion of high frequency words and lower incidence of optional *that* which are more similar to the parameters characteristic of his non-interpreted discourse. The increased frequency of fixed expressions could be then triggered by a conscious or subconscious attempt to 'conceal' the lack of confidence with more conventional language⁹ as a part of risk-avoiding strategy.

Interpreting style of the second interpreter I2 does differ significantly from his speaking style with respect to one feature and the normalised frequency values point to traces of more subtle divergence. I2 exhibits a significantly lower tendency to repetitiveness in simultaneous interpretation than in non-interpreted discourse. This feature is consistent in his interpretations from French and from German. Corpora of texts

⁹ As explained in the previous Chapter, the subjective feeling of lower confidence concerning the interpretation from Spanish, in the case of accredited interpreters with extensive professional experience cannot be considered any factual lack of competence.

interpreted by I2 from French and from German do not vary from each other with respect to repetitiveness, but each of the two differs significantly from the corpus of I2's non-interpreted discourse. No such consistency could be spotted in texts interpreted by I1, none of which deviates from his non-interpreted speech. It must be stressed that the source speeches of interpretations were read out and it is possible that the mode of delivery of the source speech might have augmented the discrepancy between I2's speaking and interpreting style. Additionally, I2's interpretations seem to point to a more conventional language use than his non-interpreted discourse. Such conclusions may be drawn judging by normalised frequency values of fixed expressions most characteristic of spoken professional English. Such a subtle and not very consistent tendency merely points to a possible less pronounced characteristic trait of the interpreter's style and should be investigated on a bigger collection of texts.

It appears that the tendency to simplification, or the lack of it, is in case of interpreting a far more complex issue than in translation. In general, the results of the present study have in most cases contradicted Laviosa's (1998) observations regarding translation and are in line with Sandrelli and Bendazolli's (2005) findings, which suggest that simplification in interpreting may be dependent on the language pair. As evident from the analyses presented in Chapter One, the tendency to use more repetitive language may depend on the source language and the mode of delivery of the source speech. The extent of this impact remains to be estimated by future research, but the outcomes of the present analysis seem to imply that the mode of delivery may exert greater effect on some language pairs than on others (as visible in Section 3.1.1.1. despite a similar proportion of texts interpreted from read speeches the list head parameters are different in the case of interpretations from French and from Dutch). None of these factors, however, have proved to significantly affect the level of informativeness or lexical sophistication of interpreted language, which have revealed to be consistently higher in all examined interpretations than in native English speeches.

Interpreted texts do not exhibit as prominent a tendency to explicitness as translations. Olohan and Baker's (2000) observations regarding the increased frequency of optional *that* in translation have been confirmed in all interpreting subcorpora of TIC, while Kruger and Van Rooy's (2010) linking adverbials and Muttasayire's (2004) apposition markers have not pointed to a uniform tendency to explicitness in interpreting. While translations demonstrate a significantly higher incidence of optional *that*, linking adverbials

and apposition markers than native English texts, interpretations manifest greater explicitness only with respect to the first parameter. Higher frequency of optional *that* after reporting verbs is indeed a sign of greater explicitness, but it may be argued, that it could just as well be a ‘side-effect’ of the interpreters’ attempts to be more formal or more conventional in the use of language. The fact that the use of the optional complementizer is the sole parameter corroborating the explicitness hypothesis in interpreting makes such an interpretation all the more plausible. Interpreters may be less inclined towards explicitness for various reasons relating mostly to the specific character of the interpreting task. The linearity constraint increases the chance that any additional linking adverbial could in the view of consecutive statements turn misapplied, shift the focus of the original message and lead to misunderstandings or hamper comprehension. Time constraint, on the other hand, may prevent interpreters from excessive use of apposition. Additionally, simultaneous interpreting due to its immediacy involves a considerably higher risk of miscomprehension and may consequently evoke more risk-averse behaviour. In other words, when uncertain, interpreters would be more cautious not to render some utterances more explicitly as implicit formulations leave more room for interpretation on the part of the target recipient. Translators, however, may ponder on a problematic fragment long enough to arrive at a satisfactory solution that they would be eager to encode as overtly as possible to increase the reader’s comprehension and facilitate the pace of processing the text.

Also, most probably due to the immediacy of the interpreting task and the ensuing higher impact of interference, the language of interpretation appears to use conventional patterns inconsistently. While common lexical bundles are indeed more frequent in interpretations, it seems, that their occurrence is dependent on the source language because the interpretations from French deviate from native English speeches less than remaining interpretations. On the other hand, the investigated fixed expressions are more frequent in interpretations from Romance languages. It may be thus concluded that not in all interpreting situations will the interpreters use conventional language. This is, on one hand, a surprising finding, considering that reaching for more conventional language i.e. structures that are retrieved more automatically, interpreters can ‘save’ their processing capacity in this component and allocate it to a different one e.g. processing numbers or memorizing a longer passage of the original text. On the other hand, it is possible that in certain language pairs the effect of interference is stronger than convention and retrieving more typical expressions in such cases would actually require greater processing capacity.

Although the results of the present study suggest that interpreting and translation differ with respect to the tendencies to simplification, explicitation or normalisation, it does not follow that the language of interpretation is completely deprived of them. The present study advanced research in this area in that it proved at which points translation and interpreting diverge. Due to numerous differences between the two modes, they may be less pronounced or perhaps manifested on different linguistic strata that have yet to be delved into (see Shlesinger 1995, Gumul 2006). Further application of corpus linguistics methods in both interpreting and translation research is bound to allow for more advanced solutions facilitating the comparison of both modes.

The term *universal* has for a long time been regarded controversial in Translation Studies (see Section 1.4.) as categorical in nature it is in the context of translation and Interpreting Studies applied to very elusive concepts. The meaning of the term suggests that whatever feature is denoted as universal, it applies to every instance of a given phenomenon and in a way implies that it is also inherent. The controversy is triggered by the fact that researchers in the discipline seem to use the term *universal* to denote features which are not clearly defined and appear to be merely recurrent. Such is also the case of the features examined in the present study, which set out to verify if the tendencies to simplification, explicitation and normalisation could be observed in all analysed interpretations i.e. whether they could be considered as interpreting universals and the outcomes prove that they are at best only recurrent in interpreting. Out of all tested parameters only the tendency to lexical sophistication certifying against simplification is a likely candidate for an interpreting universal, as such a trend can be observed in all examined corpora both in macro and micro analyses. All remaining features, even though significantly more frequent in interpretations than in native English speeches, when tested on large scale, have not been pronounced enough to be considered characteristic of interpreting in micro analysis or have been in fact disproved. The question is, whether a merely recurrent feature may be considered universal. If so, then not only greater lexical sophistication, but also higher level of informativeness and increased formality may, based on the outcomes of the present study, be put forward as putative interpreting universals. These features have been proved to occur in all interpreting corpora of TIC, regardless of the source language (see Chapter Three), but have not been manifested in all interpretations of individual interpreters (see Chapter Four). It is advocated in the present paper, however,

that a distinction be made between universal and recurrent features, as such a radical approach may facilitate future investigations of the ontology of those features.

The ontology of recurrent and universal features cannot be investigated only with corpus linguistics methods as they focus on the product of interpreting and translation process. Further comparative research of interpreting and translation, may, however, shed light on which features are more likely to be a result of more deliberate or fully subconscious strategies. The tendencies observed in translation and absent in interpretation may suggest that certain linguistic features that are today perceived as universal and inherent in the process of translation are in fact a result of consciously applied strategies most probably reinforced by training. With growing experience these strategies may become automatic, but it is possible that a different kind of training would trigger different responses.

The value of the present study for the discipline lies not only in the number of questions that it has successfully answered, but also in raising new ones that could be further investigated thus contributing to methodological advances in the examination of interpreted language. Both macro and micro analysis of the present study raise questions that should be addressed in greater detail in future interpreting research.

One of the issues that should be further explored is the effect of the mode of delivery of the original speech on universals. Present research indicates that the mode of delivery of the source speech might have an impact on repetitiveness of the target text, but this impact may vary across different language pairs. This problem could not have been sufficiently explored due to the make-up of the oral subcorpora of TIC. As only the corpus of interpretations from Spanish is balanced with this respect, simplification could be properly investigated only on Spanish-English language pair. The results point to a significant difference between texts interpreted from different modes with respect to repetitiveness, and suggest that no tendency to simplification can be observed in the analysis of the degree of informativeness or lexical sophistication. To discover if similar trends hold for explicitness and normalisation one would need a different size of corpora. An analysis, based on the frequency of lexical items, calls for sets of yet larger corpora compiled of texts interpreted from different modes that would enable a comparison of a sufficient number of such parameters as linking adverbials, apposition markers or lexical bundles.

An interesting tendency was observed also in the micro analysis of interpreting style, but due to the small scale of the study no generalisation could be made. It appears that the fact that one of the interpreters feels less confident in interpretation from a specific language may be correlated with his tendency to lower explicitness and greater normalisation. Such features could be perceived as conscious or subconscious risk avoidance strategies. This apparent trend should be inspected more scrupulously, as determining the circumstances at which tendencies to explicitness or normalisation are increased may shed light on the ontology of interpreting universals and possibly also translation universals.

It is hoped that the innovative research solutions presented in this thesis will contribute to the further methodological advancement of the branch of Translation and Interpreting Studies that investigates the nature of the translated and interpreted language. It seems that a joint perspective on universals and style has a lot to offer, as it helps distinguish potentially universal patterns not merely at a macro level, where the cumulative effect of many interpreted or translated texts may overshadow individual examples that defy the universal trend, but also at a micro level that allows to prove or disprove a hypothesised universal at the same time indicating which of the examined features are rather to be attributed to idiosyncratic stylistic preferences. Similarly, it has proved worthwhile in this particular case to examine the recurring patterns by comparing interpreting and translation corpora with respect to the source languages, as opposed to joining all the texts into two larger corpora where source languages are mixed. Such an approach makes it possible to discern features that typify all interpreted or translated texts from patterns characteristic of selected language pairs only. These and other solutions adopted in the present study may be not only be further developed and refined by others, but also serve as a spring board for new approaches that will eventually increase our understanding of the interpreted and translated language.

References

- Aijmer, Karin – Bengt Altenberg (eds.).1992. *English Corpus Linguistics*, London/New York: Longman.
- Bailey, Charles. Karl Maroldt.1977. “The French lineage of English”, in: Meisel, Jürgen, *Langues en contact – Pidgins – Creoles*. Tübingen: TBL-Verlag Narr, 21–53.
- Baker, Mona. 1993. “Corpus Linguistics and Translation Studies: Implications and Applications” in: Mona Baker – Gill Francis – Elena Tognini-Bonelli (eds.), 233-250.
- Baker, Mona. 1996 "Corpus-based Translation Studies: The Challenges that Lie Ahead", *Terminology, LSP and Translation: Studies in Language Engineering in Honour of Juan C. Sager* , Amsterdam & Philadelphia: John Benjamins.
- Baker, Mona. 2000. "Towards a methodology for investigating the style of a literary translator", *Target* 12(2), 241-266.
- Baker, Mona. 2001. *Routledge Encyclopaedia of Translation Studies*. London and New York: Routledge.
- Baker, Mona. 2004. “A Corpus-based view of similarity and difference in Translation”, *International Journal of Corpus Linguistics* 9(2): 167-193.
- Baker, Mona – Gill Francis – Elena Tognini-Bonelli (eds).1993. *Text and technology: In honour of John Sinclair*. Amsterdam | Philadelphia: John Benjamins.
- Beaton, Morven. 2007. “Interpreted Ideologies in Institutional Discourse. The Case of the European Parliament”, *The Translator* 13(2): 271-296.
- Becher, Viktor. 2011. *Explicitation and implicitation in translation. A corpus-based study of English-German and German-English translations of business texts*. (An unpublished PhD thesis. Hamburg: University of Hamburg).

- Behrens Bergljot – Cathrine Fabricius-Hansen. 2003. “On discourse coherence in translation” (A paper delivered at the Corpus-based Translation Studies: Research and Applications conference, Pretoria, South Africa, 22-25 July 2003).
- Biber, Douglas. 1995. *Dimensions of register variation: A cross-linguistic comparison*. Cambridge: Cambridge University Press.
- Blum-Kulka, Shoshana. 1986. “Shifts of cohesion and coherence in translation”. in House Juliane and Shoshana Blum-Kulka (eds.), *Interlingual and Intercultural Communication*. Tübingen, Gunter Narr. Verlag, 17-35.
- Blum-Kulka, Shoshana – Levenston. Edward. 1983. *Universals of lexical simplification*, in: Faerch, Claus – LGabriele Kasper (eds.). *Strategies in Interlanguage Communication*. London & New York: Longman. 119-140.
- Bosseaux, Charlotte. 2003. “The translator’s discursive presence: A corpus-based study”. (A paper delivered at the Corpus-based Translation Studies: Research and Applications conference, Pretoria, South Africa, 22-25 July 2003).
- British National Corpus. 2005. (<http://www.natcorp.ox.ac.uk>) (date of access: 22 Nov. 2009).
- Burnett, Scott. 1999. A corpus-based study of translational English (an unpublished MSc dissertation. Manchester: University of Manchester).
- Chesterman, Andrew. 2004. “Hypotheses about translation universals”, in: Hansen, Gyde – Kirsten Malmkjaer – Daniel Gile (eds.), *Claims, Changes and Challenges in Translation Studies*. Amsterdam: Benjamins, 1-13.
- Donaldson, Bruce. 2008. *Dutch: A comprehensive grammar*. London: Routledge.
- European Parliament Simultaneous Interpretation Corpus (EPIC). 2004. (<http://sslmitdev-online.sslmit.unibo.it>) (date of access: 20 Nov. 2009)
- Even-Zohar, Itamar 1978. “The position of Translated Literature within the Literary System”, in: James S. Holmes – Jose Lambert – Raymond van der Broeck (eds.), 117-127.
- Fontein, A.M – A. Pescher-ter Meer. 2009. *Nederlandse Grammatica voor Anderstaligen*. Utrecht : Nederlands Centrum Buitenlanders
- Gellerstam, Martin. 1986. “Translationese in Swedish Novels Translated from English”, in: Lars Wollin, Hans Lindquist (eds): *Translation Studies in Scandinavia. Proceedings from The Scandinavian Symposium on Translation Theory (SSOT)*. (*Lund Studies in English 15.*) Lund: CWK Gleerup, 88-95.

- Gumul, Ewa. 2006. "Explicitation in Simultaneous Interpreting: A strategy or a by-product of language mediation?", *Across Languages and Cultures. A Multidisciplinary Journal for Translation and Interpreting Studies* 7(2). 171-190.
- Frawley, William. 1984. *Translation. Literary, Linguistic and Philosophical Perspectives*, Newark: University of Delaware Press.
- Hermans, Theo 1996, "The translator's voice in translated narrative", *Target*, 8, 23– 48.
- Holmes, James S. 1972/1988. "The name and nature of Translation Studies", in: Holmes, James S. (ed.) , 66–80.
- Holmes, James S. 1973. "On matching and making maps: From a Translator's Notebook", *Delta* 16(4): 67-82.
- Holmes, James S. 1988. *Translated! Papers on Literary Translation and Translation Studies*, Amsterdam: Rodopi
- Holmes, James S. – Jose Lambert – Raymond van den Broeck (eds.). 1978. *Literature and Translation: New perspectives in Literary Studies*. Leuven: Acco.
- House, Juliane – Shoshana Blum-Kulka (eds). 1986. *Interlingual and Intercultural Communication*, Tübingen: Gunter Narr. Verlag.
- House, Juliane. "Cross-Cultural Pragmatics and Translation", in: Shreve, Gregory – Albrecht Neubert – Klaus Gommlich. *Kent Forum on Translation Studies* vol. 2 Basic Issues in Translation Studies. 391-403. Kent State University Press. 1996.
- House, Juliane. 2004. "Explicitness across languages", in: House, Juliane – Werner Koller – Klaus Schubert (eds.). *Neue Perspektiven in der Übersetzungs- und Dolmetschwissenschaft. BochumLAKS*, 185-208
- Johansson, Stig. 1992. "Times Change, and So Do Corpora", in: Aijmer, Karin – Bengt Altenberg, (eds.), *Corpus Linguistics* London/New York: Longman, pp. 305-314.
- Klaudy, Kinga. 1996. "Explicitation", in: Baker, M. (ed.) *Encyclopedia of Translation Studies*. (ed.) London: Routledge, 102-103.
- Klaudy, Kinga. 2001. "The Asymmetry Hypothesis. Testing the Asymmetric Relationship between Explicitations and Implicitations". Paper presented at the Third International Congress of the European Society for Translation Studies, "Claims, Changes and Challenges in Translation Studies", 30 August-1 September 2001, Copenhagen.

- Klaudy, Kinga – Krisztina Károly. 2005. “Implication in Translation: Empirical Evidence for Operational Asymmetry in Translation”. *Across Languages and Cultures* . Vol. 6. No. 1. 13–28.
- Kemppanen, Hannu. 2003. “Key words and the ideological function of translations: A corpus-based analysis”. (A paper delivered at the Corpus-based Translation Studies: Research and Applications conference, Pretoria, South Africa, 22-25 July 2003).
- Kenny, Dorothy. 1999. Norms and creativity: Lexis in translated text. (Unpublished PhD dissertation. Manchester: Centre for Translation and Intercultural Studies UMIST).
- Kenny, Dorothy. 2001a. “Corpora in Translation Studies”. in Baker, Mona (ed.), *Routledge Encyclopaedia of Translation Studies*. London and New York: Routledge, 50-53.
- Kenny, Dorothy. 2001b. *Lexis and Creativity in Translation. A Corpus-based Study*. Manchester: St. Jerome.
- Kenny, Dorothy. 2005. “Parallel Corpora and Translation Studies: old questions, new perspectives? Reporting that in Gepcolt. A case study” in: Barnbrook, Geoff, Pernilla Danielsson, Michaela Mahlberg, (eds.) *Meaningful texts: the extraction of semantic information from monolingual and multilingual corpora*. London - New York: Continuum International Publishing Group Ltd., 154-165.
- Kruger, Arlet. 2003. “CTS and Shakespeare: A South African perspective”. (Paper delivered at the Corpus-based Translation Studies: Research and Applications conference, Pretoria, South Africa, 22-25 July 2003).
- Kruger, Haidee – Bertus Van Rooy. 2010. Features of non-literary translated language: A pilot study. Proceedings of Using Corpora in Contrastive and Translation Studies, England, July (http://www.lancs.ac.uk/fass/projects/corpus/UCCTS2010Proceedings/papers/Kruger_Van_%20Rooy.pdf date of access: 13 November 2009).
- Kujamäki, Pekka – Anna Mauranen. 2004. *Translation universals: Do they exist?* Amsterdam/Philadelphia: John Benjamins.
- Leuven-Zwart, Kitty M. van – Ton Naaijken (eds.) 1991. *Translation studies: The state of the art. Approaches to Translation Studies 9*. Amsterdam/Atlanta: Rodopi.
- Laviosa-Braithwaite, Sara. 1996. The English Comparable Corpus (ECC): A Resource and a Methodology for the Empirical Study of Translation. (Unpublished PhD dissertation. Manchester: University of Manchester).

- Laviosa-Braithwaite, Sara. 2001. "Universals of translation", in: Mona Baker (ed.), 289-292.
- Laviosa, Sara. 1998. "Core patterns of lexical use in a comparable corpus of English narrative prose". *Meta* 43(4): 557-570
- Laviosa, Sara. 2002. *Corpus-based Translation Studies: Theory, findings, applications*. Amsterdam – New York: Rodopi.
- Laviosa, Sara. 2008. "Description in translation classroom: Universals as a case in point" in: Pym, Anthony – Miriam Shlesinger – Daniel Simeoni. *Beyond Descriptive Translation Studies: Investigations in homage to Gideon Toury*. John Benjamins Publishing Company: Amsterdam/ Philadelphia, 119–132.
- Lewandowska-Tomaszczyk, Barbara. 2005. *Podstawy językoznawstwa korpusowego*. Łódź: Wydawnictwo Uniwersytetu Łódzkiego.
- Liu, Dilin. 2003. "The most frequently used spoken American English idioms: A corpus analysis and its implications", *TESOL Quarterly*, 38(4), 671-700.
- Longman Dictionary of Contemporary English. 2011. (<http://www.ldoceonline.com>) (date of access: 25 November 2011).
- Malmkjær, Kirsten. 1997. "Punctuation in Hans Christian Andersen's stories and in their translations into English". in: Fernando Poyatos (ed.), 151-62.
- Masubelele, Rose. 2003. "A corpus-based appraisal of shifts in language use and translation policy in Zulu translations of the Book of Matthew (1948-1997)" (A paper delivered at the Corpus-based Translation Studies: Research and Applications conference, Pretoria, South Africa, 22-25 July 2003).
- Mauranen, Anna. 2004. "Corpora, interference and translation", in Mauranen, Anna – Pekka Kujamäki, *Translation Universals. Do they exist?*. Amsterdam: John Benjamins, 65-82.
- May, Rachel. 1997. "Sensible elocution. How translation works in and upon punctuation", *The Translator* 3(1):1-20.
- Moon, Rosamund ed. 1998. *Fixed expressions and idioms in English: a corpus-based approach*. Oxford: Clarendon Press.
- Mutesayire, Martha. 2004. "Apposition markers and explicitation: A corpus-based study", *Language Matters: Studies in the Languages of Southern Africa*. Special issue: *Corpus-based Translation Studies: Research and Applications*, 35 (1): 54-69.

- Niska, Helge. 2003. "Descriptive Translation studies: finding the norms of interpreting" (Paper delivered at the Corpus-based Translation Studies: Research and Applications conference, Pretoria, South Africa, 22-25 July 2003).
- Oakes, Michael P. 1998. *Statistics for Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Ono, Takahiro – Hitomi Tohyama – Shigeki Matsubara. 2008. Construction and Analysis of Word-level Time-aligned Simultaneous Interpretation Corpus, Proceedings of 6th International Conference on Language Resources and Evaluation (http://slp.itc.nagoya-u.ac.jp/web/papers/2008/ono_lrec2008.pdf) (date of access: 13 November 2009).
- Olohan, Maeve. 2004. *Introducing Corpora in Translation Studies*. London and New York: Routledge.
- Olohan, Maeve – Mona Baker. 2000. "Reporting *that* in Translated English: Evidence for subconscious processes of explicitation?", *Across Languages and Cultures* 1(2): 141-158.
- Øverås, Lyn. 1998 "In search of the third code: An investigation of norms in literary translation", *Meta* 43(4): 557-570.
- Papai, Vilma. 2004. "Explicitation. A universal of translated text?" in: Kujamaki Pekka – Anna Mauranen. (eds.) 144-165.
- Pokorn, Kocijančič. 2005. *Challenging the traditional axioms: Translation into a non-mother tongue*. Amsterdam: John Benjamins Publishing Company.
- Poyatos, Fernando. (ed.) 1997. *Nonverbal communication and translation*. Amsterdam - Philadelphia: John Benjamins
- Pym, Anthony. 2005. "Explaining Explicitation", in: Karoly Krisztina – Ágata Fóris (eds.) *New Trends in Translation Studies. In Honour of Kinga Klaudy*. Budapest: Akadémia Kiadó, 29-34.
- Pym, Anthony. 2007. "On Shlesinger's proposed equalizing universal for interpreting", in: Pöchhacker, Franz – Arnt Lykke Jakobsen – Inger M. Mees (eds.) *Interpreting studies and beyond: A tribute to Miriam Shlesinger*. Copenhagen Studies in Language, Samfundslitteratur Press, 175-190.
- Reynolds, Douglas A. 2008, *Gaussian Mixture Models*, Encyclopedia of Biometric Recognition, Springer.

- Saldanha, Gabriela. 2008. "Explicitation revisited: bringing the reader into the picture", *trans-kom* 1(1): 20-35.
- Saldanha, Gabriela. 2010. "Translator style: methodological considerations", *The Translator* 17(1): 25-50
- Saldanha, Gabriela. 2011. "Style of translation: the use of foreign words in translations by Margaret Jull Costa and Peter Bush", in: Alet Kruger, Kim Wallmach and Jeremy Munday (eds) *Corpus-based Translation Studies: Research and Applications*, London/New York: Continuum, 237-258.
- Sandrelli, Analisa – Claudio Bendazzoli. 2005. "Lexical patterns in simultaneous interpreting: a preliminary investigation of EPIC (European Parliament Interpreting Corpus)", *Proceedings from the Corpus Linguistics Conference Series*, 1(1), (www.corpus.bham.ac.uk/PCLC) (date of access: 15 Oct. 2009).
- Shin, Dongkwang – Paul Nation. 2008. "Beyond single words: The most frequent collocations in spoken English", *ELT Journal* 62(4):339-348
- Scott, M. Neila. 1998. *Normalisation and reader's expectations: A study of literary translation with reference to Lispector's "A Hora da Estrela"* (an unpublished PhD dissertation. Liverpool: University of Liverpool).
- Sèguinot, Candace. 1988. "Pragmatics and the Explicitation Hypothesis", *Traduction, Terminologie, Rédaction*, 1(2), 106–114. 106-113
- Shlesinger, Miriam. 1989. *Simultaneous Interpretation as a Factor in Effecting Shifts in the Position of Texts on the Oral-Literate Continuum*. (Unpublished M.A. dissertation. Tel-Aviv: Tel-Aviv University).
- Shlesinger, Miriam. 1991. "Interpreter Latitude vs. Due Process: Simultaneous and Consecutive Interpretation in Multilingual Trials", in: Sonja Tirkkonen-Condit, (ed.)147-155.
- Shlesinger, Miriam 1995. "Shifts in cohesion in simultaneous interpreting." *The Translator*. 1(2): 193-214
- Shlesinger, Miriam.1998. "Corpus-based Interpreting Studies as an offshoot of Corpus-based Translation Studies". *Meta*.43 (4), 486-493.
- Shlesinger, Miriam – Malkiel, Brenda. 2006. "Comparing Modalities: Cognates as a Case in Point". *Across Languages and Cultures* 6 (2): 173-193.
- Shlesinger, Miriam. 2009. "Towards a definition of Interpretese: An intermodal, corpus-based study", in: Hansen, Gyde, Andrew Chesterman and Heidrun Gerzymisch-

- Arbogast (eds.). *Efforts and Models in Interpreting and Translation Research: A tribute to Daniel Gile*, 237–253.
- SSLMIT. 2004. EPIC Transcription Conventions. (http://sslmitdev-online.sslmit.unibo.it/corpora/additionalpages.php?path=E.P.I.C.&source=project&content=transcription_conventions.desc) (date of access: 20 Nov. 2009).
- Somers, Harold (ed.) 1996. *Terminology, LSP and Translation: Studies in Language Engineering in Honour of Juan C. Sager*. Amsterdam & Philadelphia: John Benjamins, 175-86.
- Storms, Gotfried. 1966. "That-clauses in Modern English". *English Studies*. No. 47. 249–270.
- Stubbs, Michael. 1996. *Text and Corpus Analysis: Corpus-Assisted Studies of Language and Culture*. Oxford: Blackwell.
- Tirkkonen-Condit, Sonja (ed.) .1991. *Empirical research in Translation and Intercultural Studies*. Joensuu: University of Joensuu.
- The Translational English Corpus. 2012. (<http://www.monabaker.com/tsresources/TranslationalEnglishCorpus.htm>) (date of access: 15 March 2012).
- Toury, Gideon. 1980. *In search of a theory of translation*. Tel Aviv: The Porter Institute for Poetics and Semiotics, Tel Aviv University.
- Toury, Gideon. 1991a. "What are Descriptive Studies into translation likely to yield apart from isolated descriptions", in: Kitty M. van Leuven-Zwart – Ton Naajkens (eds.), 179-192.
- Toury, Gideon.1991b. "Experimentation in Translation Studies: Achievements, prospects and some pitfalls", in Sonja Trikkonen-Condit (ed.), 45-66.
- Toury, Gideon.1995. *Descriptive Translation Studies and Beyond*. Amsterdam and Philadelphia: John Benjamins.
- Translational English Corpus (TEC). 2003. (<http://www.monabaker.com/tsresources/TranslationalEnglishCorpus.htm>) (date of access: 20th Oct. 2009).
- Vanderauwera, Rita. 1985. *Dutch Novels Translated into English: The transformation of a Minority Literature*. Amsterdam, Rodopi.
- Venuti, Lawrence.1995. *The translator's invisibility*. London and New York: Routledge.
- Venuti, Lawrence.1998.*The scandals of translation*. London: Routledge.

- Vinay, Jean-Paul and Jean Darbelnet. 1995. *Comparative Stylistics of French and English: A Methodology for Translation, translation of Vinay and Darbelnet (1958)* by Juan C. Sager and Marie-Josée Hamel. Amsterdam and Philadelphia: John Benjamins.
- Wallmach, Kim. 2000. "Examining simultaneous interpreting norms and strategies in a South African context: A pilot corpus analysis", *Language Matters* 31: 198-221.
- Wallmach, Kim. 2003. "Legislature interpreting in Gauteng: Some insights into user expectations, interpreting strategies and norms" (A Paper delivered at the Corpus-based Translation Studies: Research and Applications conference, Pretoria, South Africa, 22-25 July 2003).
- Weinreich, Uriel. 1953. *Languages in Contact*. New York: Linguistic Circle of New York.

Appendices

APPENDIX 1

The list of function words used in the calculation of lexical density.

a, above, across, after, against, all, along, alongside, although, amid, amidst, among, amongst, an, and, any, anybody, anything, anywhere, apropos, as, at, atop, because, before, behind, below, beneath, beside, besides, between, beyond, both, but, can, can't, cos, could, couldn't, dare, daren't, despite, doesn't, don't, during, each, either, every, everybody, everyone, everywhere, except, few, for, from, he, he'd, he'll, he's, her, hers, herself, him, himself, his, how, however, if, in, inside, into, it, it'd, it's, its, itself, many, may, mayn't, me, mhm, might, mine, minus, much, must, mustn't, my, myself, needn't, neither, never, nevertheless, no, no-one, nobody, non, nonetheless, no-one, nor, not, notwithstanding, of, off, on, or, ought, oughtn't, our, ours, ourselves, out, outside, over, per, plus, shall, shan't, she, she'd, she'll, she's, should, shouldn't, since, so, some, somebody, someone, than, that, that'd, that'll, that's, the, thee, their, theirs, them, themself, themselves, then, there, there'd, there's, there've, these, they, they'd, they'll, they're, they've, thine, this, those, thou, though, through, throughout, thy, till, to, toward, towards, uhuh, under, underneath, until, up, upon, us, via, we, we'd, we'll, we're, we've, what, what'd, what's, what've, whatever, when, whenever, where, wherever, which, whichever, while, whilst, who, whom, whose, why, will, with, within, without, won't, would, wouldn't, ye, yeah, yes, yet, you, you'd, you'll, you're, you've, your, yours, yourself, yourselves, I, I'd, I'll, I'm, I've,

APPENDIX 2

The list of the 200 most frequent words in English identified by Stubbs (1996: 36-37) and used by Laviosa (1996: 119) in her investigation of translational English.

a, about, after, again, against, all, also, always, an, and, another, any, are, around, as, at, away, back, be, because, been, before, being, between, both, but, by, came, can, children, come, could, course, day, did, didn't, do, does, don't, down, each, end, er, even, every, fact, far, few, find, first, for, from, get, go, going, good, got, great, had, has, have, he, her, here, him, his, home, house, how, i, i'm, if, in, into, is, it, its, it's, just, kind, know, last, left, life, like, little, long, look, looked, made, make, man, many, may, me, mean, men, might, more, most, mr, much, must, my, never, new, no, not, nothing, now, of, off, oh, old, on, once, one, only, or, other, our, out, over, own, part, people, perhaps, place, put, quite, rather, really, right, said, same, say, says, see, she, should, so, some, something, sort, still, such, take, than, that, that's, the, their, them, then, there, these, they, thing, things, think, this, those, though, thought, three, through, time, to, too, two, under, up, us, used, very, want, was, way, we, well, went, were, what, when, where, which, while, who, why, will, with, without, work, world, would, year, years, yes, you, your

APPENDIX 3

Occurrences of optional *that* and zero-connectives after reporting verbs in TIC spoken subcorpora (raw frequencies).

	ORG_SP_EN		SI_FR_EN		SI_ES_EN		SI_DE_EN		SI_NL_EN	
	that	z.c.								
<i>admit</i>	1	0	3	0	2	0	2	0	1	0
<i>believe</i>	19	10	12	1	19	4	7	0	14	1
<i>claim</i>	1	1	0	0	0	0	1	0	0	0
<i>hope</i>	19	13	12	8	14	9	20	6	34	9
<i>know</i>	21	23	21	3	20	1	21	2	18	10
<i>say</i>	66	25	68	21	52	14	88	29	60	15
<i>suggest</i>	10	0	2	0	1	0	1	0	3	1
<i>tell</i>	9	1	4	2	2	0	4	3	4	0
TOTAL	146	73	122	35	110	28	144	40	134	36
that %	67%		78%		80%		78%		79%	

Occurrences of optional *that* and zero-connectives after reporting verbs in TIC written subcorpora (raw frequencies).

	ORG_WR_EN		TR_FR_EN		TR_ES_EN		TR_DE_EN		TR_NL_EN	
	that	z.c.	that	z.c.	that	z.c.	that	z.c.	that	z.c.
<i>admit</i>	1	0	2	0	1	0	5	0	2	0
<i>believe</i>	20	12	69	4	56	16	51	11	14	3
<i>claim</i>	1	0	2	0	0	0	2	0	1	0
<i>hope</i>	14	13	19	1	13	1	28	1	22	3
<i>know</i>	11	2	24	1	18	1	25	0	14	0
<i>say</i>	35	3	43	0	55	0	67	5	44	5
<i>suggest</i>	7	1	6	1	0	0	0	0	2	1
<i>tell</i>	8	3	24	1	6	0	6	1	3	0
TOTAL	97	34	189	8	149	18	184	18	102	12
that %	74%		96%		89%		91%		89%	

Occurrences of linking adverbials in TIC spoken subcorpora (raw frequencies).

	ORG_SP_EN	SI_FR_EN	SI_ES_EN	SI_DE_EN	SI_NL_EN
<i>consequently</i>	0	0	0	0	0
<i>as a result</i>	4	2	6	2	4
<i>hence</i>	1	1	0	0	1
<i>therefore</i>	17	26	21	43	24
<i>thus</i>	7	0	2	3	1
	29	29	29	48	30

Occurrences of linking adverbials in TIC written subcorpora (raw frequencies).

	ORG_WR_EN	TR_FR_EN	TR_ES_EN	TR_DE_EN	TR_NL_EN
<i>consequently</i>	2	3	5	1	0
<i>as a result</i>	4	3	12	15	12
<i>hence</i>	1	3	1	1	3
<i>therefore</i>	35	93	111	87	99
<i>thus</i>	2	23	6	13	18
	44	125	135	117	132

Occurrences of selected apposition markers in TIC spoken subcorpora (raw frequencies).

	ORG_SP_EN	SI_FR_EN	SI_ES_EN	SI_DE_EN	SI_NL_EN
<i>in other words</i>	1	4	3	4	11
<i>that is</i>	4	8	5	3	1
<i>that is to say</i>	0	1	4	4	0
<i>namely</i>	0	2	3	2	3
TOTAL	5	15	15	13	15

Occurrences of selected apposition markers in TIC written subcorpora (raw frequencies).

	ORG_WR_EN	TR_FR_EN	TR_ES_EN	TR_DE_EN	TR_NL_EN
<i>in other words</i>	8	5	4	7	4
<i>that is</i>	1	7	3	3	10
<i>that is to say</i>	0	1	5	1	0
<i>namely</i>	2	7	7	17	14
TOTAL	11	20	19	28	28

Lexical bundles in TIC spoken subcorpora (raw frequencies).

	ORG_SP_EN	SI_FR_EN	SI_ES_EN	SI_DE_EN	SI_NL_EN
and I think	35	26	40	53	29
at the moment	15	15	13	13	13
be able to	24	35	28	26	23
first of all	16	32	19	22	27
going to be	14	17	32	33	17
I think it	17	20	23	19	20
I think that	31	33	29	59	45
I think we	19	33	39	25	29
I would like	23	16	19	14	28
in terms of	34	32	23	30	21
is going to	12	15	21	18	18
make sure that	15	23	41	30	35
that we have	30	20	33	31	23
there is a	24	26	21	14	27
this is a	13	15	23	22	16
to have a	12	16	21	19	16
to make sure	15	16	40	25	32
we have to	41	58	109	78	69
we need to	55	82	116	72	97
would like to	31	17	26	24	37
TOTAL	476	547	716	627	622

Lexical bundles in TIC written subcorpora (raw frequencies).

	ORG_WR_EN	TR_FR_EN	TR_ES_EN	TR_DE_EN	TR_NL_EN
and it is	17	13	14	14	13
be able to	11	26	30	25	22
I do not	9	13	9	15	9
I would like	26	47	78	93	40
in order to	16	42	68	41	19
is not a	9	17	10	10	8
it is a	17	29	27	22	16
on behalf of	9	8	9	12	7
one of the	14	8	16	18	11
part of the	17	8	17	17	11
that we can	10	7	12	12	13
that we have	11	19	20	26	12
the fact that	20	15	22	34	20
there is no	18	16	13	14	15
this is a	10	19	19	18	9
to ensure that	20	28	17	21	13
we do not	20	11	27	25	7
we have to	11	10	15	17	15
we need to	25	26	47	32	32
would like to	30	39	78	94	44
TOTAL	320	401	548	560	336

Occurrences of the most frequent fixed expressions in TIC spoken subcorpora (raw frequencies).

	ORG_SP_EN	SI_FR_EN	SI_ES_EN	SI_DE_EN	SI_NL_EN
as far as	7	8	13	5	14
as to	5	10	2	6	1
as well	17	28	34	26	44
as well as	7	6	6	2	5
at all	7	5	10	9	8
first of all	16	32	19	22	27
in fact	16	16	9	19	31
in favour of	8	13	15	22	9
in order to	8	13	20	14	8
in terms of	34	32	23	30	21
kind of	7	20	52	22	13
of course	55	53	38	39	43
on behalf of	8	6	6	5	4
once again	5	11	9	13	11
sort of	7	18	12	19	15
with regard to	18	6	2	10	5
TOTAL	225	277	270	263	259

Occurrences of the most frequent fixed expressions in TIC written subcorpora (raw frequencies).

	ORG_WR_EN	TR_FR_EN	TR_ES_EN	TR_DE_EN	TR_NL_EN
as far as	6	5	5	11	14
as to	11	4	10	9	11
as well	3	2	4	4	6
as well as	13	6	10	13	8
at all	7	8	8	10	12
first of all	9	8	8	3	31
in fact	14	22	12	12	19
in favour of	6	16	9	15	12
in order to	16	42	68	41	19
in terms of	2	19	17	12	5
kind of	4	5	3	5	15
of course	31	39	29	51	43
on behalf of	9	8	9	12	7
once again	9	17	9	39	17
sort of	3	2	0	1	2
with regard to	1	18	13	33	11
TOTAL	144	221	214	271	232

Abstract in Polish

Dotychczasowa analiza tekstów tłumaczonych wykazała, iż w porównaniu z tekstami napisanymi oryginalnie w języku docelowym tłumaczenia napisane są w sposób mniej skomplikowany, jednoznacznie przedstawiają treści niejednoznaczne lub dorozumiane

w tekście źródłowym, a ich język zawiera wiele wyrażeń typowych dla danego języka docelowego. Dlatego też symplifikację, eksplicytację (lub eksplicytność) i normalizację uznano za uniwersalne cechy tłumaczenia pisemnego.

Celem badań opisanych w niniejszej rozprawie jest wykazanie, czy wymienione wyżej cechy uniwersalne przekładu pisemnego są tożsame z cechami uniwersalnymi tłumaczenia ustnego, czy też znacząco się różnią i do jakiego stopnia ich intensywność uzależniona jest od indywidualnych idiosynkratycznych preferencji stylistycznych tłumacza. Za punkt wyjścia przyjęto cztery wstępne hipotezy:

(1) Wszystkie tłumaczenia ustne niezależnie od języka źródłowego wykazują większą tendencję do symplifikacji, eksplicytności i normalizacji niż przemówienia wygłoszone oryginalnie w języku docelowym.

(2) Tendencje indywidualnych tłumaczy do symplifikacji, eksplicytności i normalizacji są niezmiennie bez względu na język źródłowy.

(3) Poszczególni tłumacze różnią się od siebie pod kątem tendencji do symplifikacji, eksplicytności i normalizacji.

(4) Styl tłumaczenia indywidualnego tłumacza jest podobny do stylu jego wypowiedzi.

W celu weryfikacji pierwszej z powyższych wstępnych hipotez przeprowadzono makroanalizę parametrów wskazujących na obecność wyżej wymienionych cech na specjalnie skompilowanym korpusie tłumaczeń przemówień wygłoszonych na sesjach plenarnych Parlamentu Europejskiego: Translation and Interpreting Corpus (TIC) – korpusie umożliwiającym porównanie wyników analiz tłumaczeń pisemnych i ustnych o wspólnych tekstach źródłowych. Badanie cech symplifikacji, eksplicytności i normalizacji na podkorpusach tłumaczeń ustnych z czterech różnych języków (francuskiego, hiszpańskiego, niemieckiego i niderlandzkiego) umożliwiło stwierdzenie, które z nich mogą występować w tłumaczeniach niezależnie od języka źródłowego, a które występują częściej, ale nie mają charakteru uniwersalnego.

Trzy kolejne hipotezy zostały zweryfikowane za pomocą mikroanalizy opartej na studium przypadku, którego najistotniejszym elementem jest porównanie transkrypcji nagrań tłumaczeń wykonanych z różnych języków źródłowych przez dwóch tłumaczy i transkrypcji nagrań ich wypowiedzi na temat związany z tematyką tłumaczeń. To badanie, z kolei, pozwoliło rozróżnić cechy będące elementami stylu tłumacza od tendencji często obserwowanych w tłumaczeniu.

Szczególnością niniejszej rozprawy jest innowacyjne połączenie badań nad cechami uniwersalnymi języka tłumaczeń z analizą stylu indywidualnych tłumaczy oraz porównawcza analiza tłumaczeń z różnych języków źródłowych. Takie podejście pozwala określić, które z badanych cech mogą mieć charakter uniwersalny, które są jedynie częstym zjawiskiem, a które są w znacznym stopniu uzależnione od stylu tłumacza.