Using the features of translated language to investigate translation expertise: a corpus-based study

K.R. Redelinghuys
B.A., B.A. Hons.
20393172

Dissertation submitted in fulfilment of the requirements
for the degree Magister Artium
at the Vaal Triangle Campus of the North-West University

Supervisor: Prof. H. Kruger
Co-supervisor: Prof. A.J. van Rooy

May 2013

ACKNOWLEDGEMENTS

I would like to thank the following:

- My supervisor, Prof. Haidee Kruger, for her mentorship, enlightening guidance, patience, support and encouragement throughout this study. Without her this study would not have been possible.
- My co-supervisor, Prof. Bertus van Rooy, for his guidance and patience, and helping me with his statistical expertise.
- Jeffrey for his never-ending love and encouragement, and for making believe that I can do anything I put my mind to.
- My family and my grandmother for their unconditional love and support throughout this study.
- Val for her true friendship and for always believing in me.
- Everyone at the School of Languages who took an interest in my study and offered me advice throughout my study.
- All of the universities that sent translations to use in my corpus.

ABSTRACT

Keywords

Translation expertise, features of translated language, translation universals, corpus-based approach, descriptive translation studies (DTS)

Research based on translation expertise, which is also sometimes referred to as translation competence, has been a growing area of investigation in translation studies. These studies have not only focused on how translation expertise may be conceptualised and defined, but also on how this expertise is acquired and developed by translators. One of the key observations that arise from an overview of current research in the field of translation expertise is the prevalence of process-oriented methodologies in the field, with product-oriented methodologies used comparatively infrequently. This study is based on the assumption that product-oriented methodologies, and specifically the corpus-based approach, may provide new insights into translation expertise. The study therefore sets out to address the lack of comprehensive and systematic corpus-based analyses of translation expertise.

One of the foremost concerns of corpus-based translation studies has been the investigation of what is known as the features of translated language which are often categorised as: explicitation, simplification, normalisation and levelling-out. The main objective of this study is to investigate the hypothesis that the features of translated language can be taken as an index of translation expertise. The hypothesis is founded on the premise that if the features of translated language are considered to be the textual traces of translation strategies, then the different translation strategies associated with different levels of translation expertise will be reflected in different frequencies and distributions of these features of translated language in the work of experienced and inexperienced translators. The study therefore aimed to determine if there are significant differences in the frequency and distribution of the features of translated language in the work of experienced and inexperienced translators. As background to this main research question, the study also investigated a secondary hypothesis in which translated language demonstrates unique features that are the consequence of various aspects of the translation process.

A custom-built comparable English corpus was used for the study, comprising three subcorpora: translations by experienced translators, translations by inexperienced translators, and non-translations. A selection of linguistic operationalisations was chosen for each of the four features of translated language. The differences in the frequency and distribution of these linguistic operationalisations in the three subcorpora were analysed by means of parametric or non-parametric ANOVA.

The findings of the study provide some support for both hypotheses. In terms of the translation expertise hypothesis, some of the features of translated language demonstrate significantly different frequencies in the work of experienced translators compared to the work of inexperienced translators. It was found that experienced translators are less explicit in terms of: formal completeness, simplify less frequently because they use a more varied vocabulary, use longer sentences and have a lower readability index score on their translations, and use contractions more frequently, which signals that they normalise less than inexperienced translators. However, experienced translators also use neologisms and loanwords less frequently than inexperienced translators, which is suggestive of normalisation occurring more often in the work of experienced translators when it comes to lexical creativity. These linguistic differences are taken as indicative of the different translation strategies used by the two groups of translators. It is believed that the differences are primarily caused by variations in experienced and inexperienced translators' sensitivity to translation norms, their awareness of written language conventions, their language competence (which involves syntactic, morphological and vocabulary knowledge), and their sensitivity to register.

Furthermore, it was also found that there are indeed significant differences between translated and non-translated language, which also provides support for the second hypothesis investigated in this study. Translators explicitate more frequently than non-translators in terms of formal completeness, tend to have a less extensive vocabulary, tend to raise the overall formality of their translations, and produce texts that are less creative and more conformist than non-translators' texts. However, statistical support is lacking for the hypothesis that translators explicitate more at the propositional level than original text producers do, as well as for the hypothesis that translators are inclined to use a more neutral middle register.

OPSOMMING

Sleutelwoorde

Vertaalkundigheid, vertaalde taalkenmerke, vertaaluniversele kenmerke, korpusgebaseerde benadering, beskrywende vertaalkunde (DTS)

Vertaalkundigheid, oftewel vertaalvaardigheid, is ŉ toenemende gebied van ondersoek in die veld van vertaalkunde. Navorsingstudies hieroor fokus nie net op hoe vertaalkundigheid gekonseptualiseer en gedefinieer kan word nie, maar ook op hoe vertalers hierdie kundigheid verwerf en ontwikkel. ŉ Belangrike waarneming wat uit ŉ oorsig van huidige navorsing in die gebied van vertaalkundigheid voortspruit, is die algemeenheid waarmee prosesgeoriënteerde metodologieë voorkom en die seldsaamheid waarmee produkgeoriënteerde metodologieë gebruik word. Hierdie studie is gegrond op die aanname dat produkgeoriënteerde metodologieë, meer spesifiek die korpusgebaseerde benadering, nuwe insig in vertaalkundigheid kan bied. Dié studie onderneem dus om die gebrek aan ŉ omvattende en sistematiese korpusgebaseerde analise van vertaalkundigheid reg te stel.

Een van die fokuspunte van korpusgebaseerde vertaalkunde is die ondersoek na sogenaamde kenmerke van vertaalde taal, wat dikwels gekategoriseer as eksplisitering, vereenvoudiging, normalisering en gelykmaking. Die hoofdoelwit van hierdie studie is om die hipotese te ondersoek dat die kenmerke van vertaalde taal ŉ aanduiding van vertaalkundigheid is. Hierdie hipotese is gegrond op die veronderstelling dat, indien die kenmerke van vertaalde taal beskou word as die tekstuele spore van vertaalstrategieë, die verskillende vertaalstrategieë wat met die verskillende vlakke van vertaalkundigheid geassosieer word in die frekwensie en verspreiding van die kenmerke van vertaalde taal in die werk van ervare en onervare vertalers weerspieël sal word. Hierdie studie stel dit dus ten doel om te bepaal of daar enige beduidende verskille in die frekwensie en verspreiding van die kenmerke van vertaalde taal in die werk van ervare en onervare vertalers is. Die studie ondersoek ook as agtergrond tot die hoofnavorsingsvraag die hipotese dat vertaalde taal unieke eienskappe toon vanweë verskeie aspekte van die vertaalproses.

Die studie maak gebruik van ’n selfsaamgestelde vergelykende Engelse korpus wat uit drie subkorpora bestaan, naamlik: vertalings deur ervare vertalers, vertalings deur onervare vertalers en nie-vertalings. ’n Verskeidenheid linguistiese operasionaliserings is uit elk van die vier kenmerke van vertaalde taal gekies. Die verskille in die frekwensie en verspreiding van hierdie linguistiese operasionaliserings in die drie subkorpora is deur middel van parametriese of nieparametriese ANOVA’s geanaliseer.

Die bevindinge van die studie bied ondersteuning vir beide hipoteses. In terme van die vertaalkundigheid-hipotese, is beduidende verskille in die frekwensie van sommige van die kenmerke van vertaalde taal in die werk van ervare en onervare vertalers waargeneem. Daar is gevind dat ervare vertalers minder eksplisiet is ten opsigte van formele volledigheid; nie so gereeld vereenvoudig nie omdat hulle ’n veelkleurige woordeskat het, langer sinne gebruik en ’n laer leesbaarheidsindekstelling op hulle vertalings het; en dat hulle meer gereeld sametrekkings gebruik, wat ’n aanduiding is dat hulle minder as onervare vertalers normaliseer. Ervare vertalers gebruik egter neologismes en leenwoorde minder gereeld, wat aandui dat hulle meer as onervare vertalers normaliseer wat leksikale kreatiwiteit betref. Hierdie linguistiese verskille word beskou as ’n aanduiding van die verskillende vertaalstrategieë wat ervare en onervare vertalers gebruik en word hoofsaaklik toegeskryf aan variasies in die twee groepe se sensitiwiteit vir vertaalnorme, bewustheid van geskrewe taalkonvensies, taalvaardigheid (kennis ten opsigte van sintaksis, morfologie en woordeskat), en sensitiwiteit vir register. Daarbenewens is daar ook gevind, ter ondersteuning van die tweede hipotese wat hierdie studie ondersoek het, dat daar wel beduidende verskille tussen vertaalde en nie-vertaalde taal is. Vertalers eksplisiteer meer gereeld as nie-vertalers ten opsigte van formele volledigheid, is geneig om ’n meer beperkte woordeskat te besit, is geneig om vertalings op ’n meer formele vlak as die oorspronklike te skryf, en produseer tekste wat minder kreatief en meer konformisties is as dié van nie-vertalers. Daar ontbreek egter statistiese ondersteuning vir die hipotese dat vertalers meer op die proposisionele vlak as die oorspronklike teksskepper eksplisiteer, asook vir die hipotese dat vertalers geneig is om ’n meer neutrale middelregister te gebruik.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

1.1 Contextualisation 1
1.2 Problem statement 4
1.3 Research questions 6
1.4 Research aims 6
1.5 Hypotheses 7
1.6 Methodology 7
1.6.1 General methodological approach 7
1.6.2 Corpus analysis 8
1.6.2.1 Corpus composition 8
1.6.2.2 Data collection 8
1.6.2.3 Data processing: quantitative and qualitative analysis 9
1.7 Conclusion 10

CHAPTER 2: THEORETICAL FRAMEWORK

2.1 Introduction 11
2.2 Translation expertise 11
2.2.1 Introduction 11
2.2.2 Theoretical approaches to translation expertise 12
2.2.2.1 The componential approach to translation expertise 12
2.2.2.2 Approaches to translation expertise drawn from cognitive psychology 17
2.2.2.3 Translation expertise from the perspective of relevance theory 21
2.2.3 Implications of the three approaches to translation expertise 23
2.2.4 Differences between experienced and inexperienced translators 23
2.2.5 Methodological implications of existing studies 26
2.3 The corpus-based approach and the features of translated language 28
2.3.1 Introduction: corpora and descriptive translation studies 28
2.3.2 Universals, laws and features: recurrent patterns of translated language 31
2.3.3 The features of translated language 33

2.3.3.1 Explicitation 33
(a) Overview 33
(b) Results of previous studies 37
2.3.3.2 Simplification 41
(a) Overview 41
(b) Results of previous studies 44
2.3.3.3 Normalisation 46
(a) Overview 46
(b) Results of previous studies 48
2.3.3.4 Levelling-out 50
(a) Overview 50
(b) Results of previous studies 50
2.4 Using the features of translated language as an index of translation expertise 51
2.5 Conclusion 52

CHAPTER 3: METHODOLOGY

3.1 Introduction 53
3.2 Research design 53
3.3 Corpus composition 55
3.3.1 Compilation 55
3.3.2 Registers, word count and source languages 57
3.4 Data collection 68
3.4.1 Part-of-speech tagging 68
3.4.2 Data collection 68
3.5 Operationalisations 69
3.5.1 Explicitation 69
3.5.1.1 *That*-omission (E1 operationalisation) 69
3.5.1.2 Conjunctive markers (E2 operationalisation) 71
3.5.2 Simplification 76
3.5.2.1 Type-token ratio (S1 operationalisation) 76
3.5.2.2 Readability score (S2 operationalisation) 76
3.5.3 Normalisation 78

3.5.3.1 Contraction ratio (N1 operationalisation) 78
3.5.3.2 Neologisms and loanwords (N2 operationalisation) 80
3.5.4 Levelling-out 81
3.5.4.1 Register variation (L1 operationalisation) 81
3.6 Hypotheses 82
3.6.1 Explicitation 82
3.6.2 Simplification 83
3.6.4 Levelling-out 88
3.7 Data analysis 89
3.8 Conclusion 91

CHAPTER 4: FINDINGS AND DISCUSSION

4.1 Introduction 92
4.2 Findings 92
4.2.1 *That*-omission (E1 operationalisation) 92
4.2.2 Conjunctive markers (E2 operationalisation) 95
4.2.2.1 Elaboration 95
4.2.2.2 Extension 97
4.2.2.3 Enhancement 99
4.2.2.4 All categories of conjunctive markers 101
4.2.3 Type-token ratio (S1 operationalisation) 103
4.2.4 Readability score (S2 operationalisation) 106
4.2.5 Word length (S3 operationalisation) 108
4.2.6 Contraction ratio (N1 operationalisation) 110
4.2.6.1 Verb contractions 111
4.2.6.2 *Not*-negation contractions 113
4.2.6.3 All contracted forms 115
4.2.7 Neologisms and loanwords (N2 operationalisation) 117
4.2.8 Register variation (L1 operationalisation) 120
4.2.8.1 *That*-omission (E1 operationalisation) 120
4.2.8.2 Conjunctive markers (E2 operationalisation) 121
4.2.8.3 Type-token ratio (S1 operationalisation) 123

4.2.8.4 Readability score (S2 operationalisation) 126
4.2.8.5 Word length (S3 operationalisation) 127
4.2.8.6 Contraction ratio (N1 operationalisation) 128
4.2.8.7 Neologisms and loanwords (N2 operationalisation) 129
4.3 Interpretation of findings 130
4.3.1 Explicitation 130
4.3.2 Simplification 134
4.3.3 Normalisation 137
4.3.4 Levelling-out 140
4.4 Conclusion 141

CHAPTER 5: SUMMARY AND CONCLUSION

5.1 Introduction 143
5.2 The features of translated language in translations and non-translations 143
5.3 The features of translated language and translation expertise 145
5.4 Implications of the findings, limitations of the study, and future studies 148
5.5 Conclusion 149

REFERENCES 150

LIST OF TABLES

Table 1.1: Operationalisations of feature parameters 9
Table 3.1: Word counts by register in the three subcorpora 58
Table 3.2: Representation of varieties of English in the NT subcorpus for the five registers 59
Table 3.3: Representation of source languages for the two translation subcorpora 61
Table 3.4: Representation of source languages in the two translation subcorpora for the academic register 62
Table 3.5: Representation of source languages in the two translation subcorpora for the creative writing register 63
Table 3.6: Representation of source languages in the two translation subcorpora for the instructional register 64
Table 3.7: Representation of source languages in the two translation subcorpora for the popular writing register 65
Table 3.8: Representation of source languages in the two translation subcorpora for the reportage register 67
Table 3.9: List of reporting verbs 70
Table 3.10: List of elaboration markers 72
Table 3.11: List of conjunctive markers of extension 73
Table 3.12: List of enhancement conjunctions 75
Table 3.13: List of contractions and full forms 80

LIST OF FIGURES

Figure 2.1: Diagrammatic representation of the compential approach to translation expertise 13
Figure 2.2: Knowledge systems that form part of translation expertise 18
Figure 2.3: The task-dependent flexibility of translation expertise 20
Figure 2.4: Differences in information encoding for broadband and narrow-band translators 22
Figure 4.1: Non-normal distribution of values for the variable *that*-omission ratio 93
Figure 4.2: Boxplot of *that*-omission ratio for the three subcorpora 94
Figure 4.3: Non-normal distribution of values for the variable conjunctive markers indicating elaboration 96
Figure 4.4: Boxplot of elaboration markers for the three subcorpora 97
Figure 4.5: Normal distribution of values for the variable conjunction markers indicating extension 98
Figure 4.6: One-way ANOVA for conjunctive markers indicating extension in the three subcorpora 99
Figure 4.7: Non-normal distribution of values for the variable conjunctive markers indicating enhancement 100
Figure 4.8: Boxplot of enhancement markers in the three subcorpora 101
Figure 4.9: Normal distribution of values for the variable all conjunctive markers 102
Figure 4.10: One-way ANOVA for all conjunctive markers in the three subcorpora 103
Figure 4.11: Non-normal distribution of values for the variable standardised TTR 104
Figure 4.12: Boxplot of standardised TTR in the three subcorpora 105
Figure 4.13: Normal distribution of values for variable Flesch Reading Ease score 107
Figure 4.14: ANOVA for Flesch Reading Ease score in the three subcorpora 108
Figure 4.15: Normal distribution of values for the variable word length 109
Figure 4.16: ANOVA for mean word length in the three subcorpora 110
Figure 4.17: Non-normal distribution of values for the variable verb contraction ratio 111
Figure 4.18: Boxplot of verb contraction ratio in the three subcorpora 112
Figure 4.19: Non-normal distribution of values for the variable *not*-negation contraction ratio 113
Figure 4.20: Boxplot of *not*-negation ratio in the three subcorpora 114
Figure 4.21: Non-normal distribution of values for variable contraction ratio 115
Figure 4.22: Boxplot for overall contraction ratio 116
Figure 4.23: Non-normal distribution of values for the variable neologisms and loanwords 118
Figure 4.24: Boxplot for neologisms and loanwords 119

Figure 4.25: ANOVA for *that*-omission ratio in the three corpora and five registers 121
Figure 4.26: ANOVA for all conjunctives in the three subcorpora and five registers 123
Figure 4.27: Box-Cox transformation of data for standardised TTR 124
Figure 4.28: ANOVA for standardised TTR in the three subcorpora and five registers, using transformed data 125
Figure 4.29: ANOVA for Flesch Reading Ease score in the three subcorpora and five registers 126
Figure 4.31: ANOVA for contraction ratio in the three subcorpora and five registers 129
Figure 4.32: ANOVA for neologisms and loanwords in the three subcorpora and five registers 130

CHAPTER 1: INTRODUCTION

1.1 Contextualisation

The translation strategies, or the unconscious or conscious, non-verbal or verbal procedures (Molina & Albir, 2002:508), used by experienced and inexperienced translators[1] have been investigated by a number of translation researchers, who aim to describe and explain the differences that are evident between the two sets of translators from process- or product-oriented viewpoints (see Jääskeläinen, 1993; Künzli, 2004; Lörscher, 2005). The concept of translation expertise,[2] also referred to as translation competence by some theorists (Albir & Alves, 2009:63), is of key importance in such studies. Even though it has proved difficult to provide an adequate or acceptable definition of what translation expertise or competence entails (Pym, 2003), it may be assumed, minimally, that translation expertise consists of the things professional translators are able to do in terms of skills, what they know in the sense of knowledge, and their relation towards others in terms of dispositions or attitudes (Pym, 2010).

Pym (2010) believes that translation expertise can be reduced to two components, namely declarative knowledge, which Anderson (1982:370) defines as the acquisition of data and rules; and operational knowledge, which is the application of the specific data and rules to resolve problems (Anderson, 1983:215) – a view that is supported by Alves and Gonçalves (2007:42) and Munday (2009:235). Some studies of translation expertise have found that one of the important distinctions in the ways in which experienced and inexperienced translators translate is related to the development of their different kinds of knowledge. Experienced translators, on the one hand, are able to apprehend the

[1] In existing studies, there is significant variation in how "professional" or "expert translators" and "novice translators" are defined (Jääskeläinen, 2010; Pym, 2010). Harris (2010), for instance, defines "expert translators" as people who have undergone some translation training, while other theorists define "novice translators" as translators who have received training up to the MA level (Pym, 2010). To avoid terminological difficulty, this study defines "experienced translators" as translators for whom translation is a principal professional activity (PACTE, 2005). "Inexperienced translators" are defined broadly to include student translators as well as laypersons doing translation, comprising translators with little or "no previous experience in translation proper" (Jääskeläinen, 1993:99).

[2] In this study, the term "translation expertise" will be used rather than the term "translation competence", to avoid associations with the Chomskyan distinction between competence and performance (see Pym, 2010). Another reason for this preference is based on the observation of Shreve (2002:154) that translation graduates may exhibit varying degrees of translation competence, as they have been taught at their respective translation schools, but not translation expertise. The most important factor influencing translation expertise is "deliberate practice in the domain" – a requisite which translation graduates will lack, but which a practising translator with a number of years of experience will have (Shreve, 2002:154). The term "translation expertise", in other words, evokes the notion of experience, which is not necessarily the case for the term "translation competence".

more complex or deeper structures of the source text as they are able to see the translation task on an abstract level or see the text as an organic whole, as a consequence of the fact that they have acquired both declarative knowledge and operational knowledge (Shreve, 2002:163). Inexperienced translators, on the other hand, rely mainly on declarative knowledge, and for this reason they tend to focus on the translation problem's superficial aspects, such as the linguistic structures of the source text (Shreve, 2002:164).

Despite the fact that research such as that of Kiraly (1995), Gutt (2000) and PACTE (2011) has provided valuable information about translation expertise in general, the field of enquiry is not without its limitations. In the context of this study, one of the most important limitations identified is the fact that even though translation expertise can be studied utilising both process- and product-oriented methodologies, the majority of studies have made use of the former, with a particular focus on the use of think-aloud protocols (TAPs) (PACTE, 2002:42). Product-oriented methodologies appear to be much less frequently used as a method for investigating translation expertise, even though some scholars, such as Alves *et al*. (2010) and Alves and Vale (2011) argue the importance of integrating corpora in process-oriented studies of translation. In this regard, it is particularly interesting to note that the concept of translation expertise has not frequently been linked to the burgeoning field of corpus-based translation studies, which reflects a product-based methodology used to investigate various aspects of translated language (see Olohan, 2004; Zanettin, 2012 for overviews of the field).

One of the uses corpus-based translation studies has been applied to is the investigation of what is known as translation universals (Baker, 1996) or, in less totalising terms, as the features of translated language (Olohan, 2004:93).[3] The features of translated language can be defined as the "linguistic features which typically occur in translated rather than original texts and are thought to be independent of the influence of the specific language pairs involved in the process of interpretation" (Laviosa, 2008a:306). It has been proposed that these features include explicitation, simplification, normalisation and levelling-out (Baker, 1996), and interference (Toury, 2012; Tirkkonen-Condit, 2002; Mauranen, 2004). Each of these features has been operationalised in various ways in existing

[3] There is some contention about the term "translation universals" as it suggests that these translation patterns will be observed irrespective of the language, period, genre or translator (Kenny, 2001:53). However, as Kenny (2001:53) points out, it is difficult to distinguish between translation features that are a result of culture-specific translation norms and those that truly represent universals. Since this study is not primarily concerned with the debate surrounding universalism, the term "features of translated language" will be used rather than "universals of translation".

research on the features of translated language. In this study, eight operationalisations for the four features investigated (explicitation, simplification, normalisation and levelling-out)[4] were selected. These operationalisations are outlined in more detail in Section 1.6.2.2, and discussed in detail in Chapter 3.

Most studies investigating the features of translated language (see, for example, Laviosa, 1998b; Mutesayire, 2004; Yuan & Gao, 2008; Kruger & Van Rooy, 2012) have focused on translations produced by experienced, professional translators (Castagnoli, 2008) as prototypical instances of the category "translation". There has been some research on the features of translated language evident in the translations of experienced and inexperienced translators. However, as will become evident from the discussion below, these studies are not only limited in number, but have also frequently produced conflicting results.

Explicitation is defined as "an overall tendency to spell things out rather than leave them implicit in translation" (Baker, 1996:180). Blum-Kulka (1986:293) suggests that explicitation will be more pronounced in the translations of inexperienced translators, as "the less experienced the translator, the more his or her process of interpretation of the SL might be reflected in the TL". However, she notes that explicitation is also practiced by professional or experienced translators (Blum-Kulka, 1986:294), which suggests a difference of degree (rather than an absolute difference) between the translations of experienced and inexperienced translators, as with the other features of translated language. However, a study by Englund Dimitrova (2003) reveals that student or inexperienced translators tend not to add connectives in translations from Russian to Swedish, whereas experienced translators tend to explicitate the relationships between ideas by utilising connectives.

Studies of simplification, or "the tendency to simplify language used in translation" (Baker, 1996:181), in the context of translation expertise have also yielded inconclusive results. Pastor *et al*. (2008) have attempted to compare the extent to which experienced and inexperienced translators simplify their translations. Using measurements of sentence length and analysing the use of simple or complex sentences, they found that translations by inexperienced translators do not, overall, possess more simplification traits compared to translations produced by experienced translators (Pastor *et al*.,

[4] The comparable corpus design (see section 1.6.2.1, section 2.3.1, and Chapter 3) is not well suited to the investigation of interference; for investigating this feature a parallel design would be a better choice. For this reason, interference is not investigated in this study.

2008:7). However, their findings are limited by the fact that their corpus consisted only of medical translations and comparable original medical texts (Pastor *et al*., 2008:2), and did not include texts of other genres.

In terms of normalisation or conservatism, "the tendency to exaggerate features of the target language and to conform to its typical patterns" (Baker, 1996:183), it appears that existing studies have mainly focused on translations done by experienced, professional translators (see May, 1997; Kenny, 1998; Kenny, 2001; Williams, 2005; Yuan & Gao, 2008), as have studies on levelling-out, or the tendency of translated language to gravitate towards the centre of any continuum (see Laviosa, 1998b).

Lastly, as far as source-text or -language interference is concerned, Kujamäki (2004) studied the under-representation of target language specific elements (or "unique items") in translations produced by inexperienced translators. The "unique items" hypothesis postulates that translations have a universal tendency to "manifest lower frequencies of linguistic elements that lack linguistic counterparts in the source languages" (Tirkkonen-Condit, 2002:209) in terms of lexis, syntax, phraseology, collocational patterns, textual organisation and pragmatic functions (Tirkkonen-Condit, 2002:216). This hypothesis is related to the translation feature of interference (Kujamäki, 2004:198; Tirkkonen-Condit, 2004:182) in the sense that the linguistic features that pertain to the composition of the source text are transferred to the target text (Laviosa, 2008a:308) rather than being translated with elements or features specific to the target language. In his study of this hypothesis, in the context of translation expertise, Kujamäki (2004:192) found that inexperienced translators tend to prefer solutions which are motivated directly by the lexical surface structures of the source texts translated.

1.2 Problem statement

From the above it emerges that corpus-based approaches have not been used to any significant degree to investigate the concept of translation expertise. This study aims to address the lack of comprehensive and systematic corpus-based investigations of translation expertise, utilising the hypothesised features of translated language as an indicator of translation expertise so as to gain more information about the phenomenon from a descriptive perspective. The study is comprehensive owing to the investigation of the co-occurrence of features, in order to acquire a broader picture of the differences between translations produced by experienced and inexperienced translators. Using comprehensive product-based data may make a contribution towards a more nuanced understanding

of a concept (translation expertise) that has, up to now, largely been investigated by means of process studies.

In view of the discussion above, it is proposed that the features of translated language will not only demonstrate significantly different frequencies when translations are compared to non-translations, but also that there will be noticeable differences in this respect between the translations of experienced and inexperienced translators. The basic assumption is that the extent to which the features of translated language, such as explicitation, simplification, normalisation and levelling-out, occur may be seen as a reflection of the different translation strategies used by experienced and inexperienced translators. These differences may be the result of various (interrelated) processes. According to Laviosa (2008a:307), the features of translated language may be the consequence of the communicative nature of translation, translators' awareness of their socio-cultural roles, and of the constrained bilingual cognitive processing involved in translation. In general, it can be expected that experienced translators will be more aware of the communicative role of translation and of the socio-cultural constraints within which they operate, in comparison to inexperienced translators. Cognitive development will also be a contributing factor as experienced translators should be able to use a combination of declarative and operational knowledge to solve translational problems, whereas inexperienced translators would mainly rely on declarative knowledge.

On the other hand, some of the features of translated language may be expected to occur in translations produced by inexperienced translators precisely because of their lack of experience. This is caused by the fact that inexperienced translators often have a micro-level approach to translation in that their translation process is usually limited to the lexical level of texts. Inexperienced translators tend to focus on the individual words or signs of the source language, which means that the source text is attributed with a dominant role (Lörscher, 2009:13). As a result, it may be expected that translations produced by inexperienced translators will be characterised by distinctive linguistic regularities or patterns which will differ from those of non-translated texts as well as from those of translated texts produced by experienced translators.

Clearly, the exact effects of (a) the translation process and (b) the level of expertise of the translator are interwoven in complex ways. This dissertation is, broadly, an investigation of the hypothesis that the features of translated language may be used as indicators of the linguistic regularities that

distinguish originally produced texts from translated texts, as well as translations produced by experienced translators from translations produced by inexperienced translators. In the latter case, the argument is particularly interested in the possibility of viewing these differences in linguistic regularities as a reflection of the different translation strategies and different kinds of skills of experienced and inexperienced translators.

1.3 Research questions

The discussion above leads to the following research questions:

1. What are the differences in the frequency and distribution of the features of translated language in translations and non-translations?

2. What are the differences in the frequency and distribution of the features of translated language in the translations of experienced and inexperienced translators?

3. How may the differences in the frequency and distribution of the features of translated language in the work of experienced and inexperienced translators be linked to differences in levels of experience, particularly as far as cognitive processing in translation, and role perception are concerned?

1.4 Research aims

The aims of this study are to:

1. Identify the differences in the frequency and distribution of the features of translated language in translations and non-translations.

2. Identify the differences in the frequency and distribution of the features of translated language in the translations of experienced and inexperienced translators.

3. Explain how the differences in the frequency and distribution of the features of translated language in the work of experienced and inexperienced translators may be linked to differences

in levels of experience, particularly as far as cognitive processing in translation, and role perception are concerned.

1.5 Hypotheses

It is hypothesised that:

1. Significant differences will be evident in the frequency and distribution of the features of translated language between translations and non-translations for the operationalisations associated with explicitation, simplification, normalisation and levelling-out.

2. Significant differences will be evident in the frequency and distribution of the features of translated language between the translations of experienced and inexperienced translators for the operationalisations associated with explicitation, simplification, normalisation and levelling-out.

3. The reasons for these differences may be ascribed to at least three factors: a greater awareness of the socio-cultural role of translators, the communicative function of translation on the part of experienced translators, as well as more developed cognitive translation skills in this group.

1.6 Methodology

1.6.1 General methodological approach

In order to answer the research questions and test the hypotheses set out above, the study utilises a corpus-based approach, with a combination of quantitative and qualitative analysis. Corpus-based translation studies is informed by a linguistic area known as corpus linguistics that entails the study of large computerised corpora of texts with computer programmes (Kruger, 2002:70). This type of research in translation studies is concerned with, among other research questions, the identification of recurrent features of translation, by means of a combination of hypotheses and theoretical concepts which are applicable to both product- and process-oriented research, and inductive and deductive research (Kruger, 2002:70).

1.6.2 Corpus analysis

1.6.2.1 Corpus composition

The study is carried out using a comparable corpus of English texts, consisting of two distinct text collections of translations into English, as well as a reference subcorpus comprising non-translational texts in the same language. The one text collection of the monolingual comparable corpus is made up of translations produced by experienced translators, whereas the other collection consists of translations done by inexperienced translators. The reference subcorpus contains a collection of original English writing and therefore only consists of non-translations. It is thus possible to carry out a three-way comparison to determine whether there are significant differences in the frequency of the features of translated language in non-translated texts, texts produced by experienced translators, and texts produced by inexperienced translators. This comparable approach was used rather than a parallel-corpus approach (which consists of source-language texts and their translations) as the former is generally the methodology followed in corpus-based studies of the features of translated language. Furthermore, since this study does not explicitly investigate features related to source-text and source-language interference, no comparative analysis of source and target texts is necessary.

The three subcorpora were constructed to be as comparable as possible, in terms of size, genre representation, historical timeframe of production, and other relevant parameters, to aim for and ensure a balanced representation across the three subcorpora. In addition, five different registers are represented in the corpus, and for the translation subcorpora, an attempt was made to include translations from a number of different source languages. Care was taken with the construction of the corpus so as to address some of the shortcomings of existing studies.

1.6.2.2 Data collection

Data were collected by using various functions in WordSmith Tools 5.0 (Scott, 2008) and a combination of manual and automatic classification. The four feature parameters selected for investigation are explicitation, normalisation, simplification and levelling-out. Each of these four parameters was operationalised in a number of ways, based on linguistic elements and structures linked to the feature parameters evident in existing research. The operationalisations selected for each parameter are set out in Table 1.1.

Feature parameter	Operationalisation code	Operationalisation description
Explicitation	• E1 • E2	• Omission ratio of reporting *that* • Frequency of conjunctive markers
Simplification	• S1 • S2 • S3	• Standardised type-token ratio • Mean readability score • Mean word length
Normalisation	• N1 • N2	• Contraction ratio • Neologisms and loanwords
Levelling-out	• L1	• Register variation

Table 1.1: Operationalisations of feature parameters

1.6.2.3 Data processing: quantitative and qualitative analysis

The analysis of the data primarily made use of quantitative methods, with some use of qualitative analysis. As far as quantitative statistical methods are concerned, Analysis of Variance (ANOVA) was used as the primary method of analysis. ANOVA is a procedure which tests for significant statistical differences that occur between means for variables or groups (Hill & Lewicki, 2006:43). One of the main reasons why ANOVA was used as a method is that, compared to some other statistical methods for testing for differences between variables or groups (such as the t-test), it is more powerful in the sense that each factor can be tested while the others are controlled for – which also means that fewer observations are required for a significant effect to be located (Hill & Lewicki, 2006:45).

In this study, using ANOVA therefore made it possible to investigate the effects of multiple interacting factors, including the effects related to translation versus original writing, translation expertise, and register. ANOVA analyses variance in that it divides the overall variance into the component that is, as a result of random error, and the components that are, as a result of differences between means (Hill & Lewicki, 2006:44). The latter components of variance are, in turn, tested for statistical significance. If they are found to be significant, the null hypothesis (H_0) of no difference between means is rejected, and the alternative hypothesis (H_a) – that the means differ from one another – is accepted (Hill & Lewicki, 2006:44).

One of the conditions for using ANOVA is normal distribution of data. In instances where data were not normally distributed, the non-parametric equivalent of ANOVA, the Kruskal-Wallis ANOVA, was used to test for significant differences between the three subcorpora. The non-parametric Kruskal-Wallis ANOVA is essentially similar to the parametric ANOVA but, unlike the parametric ANOVA, it does not require or assume a normal distribution of the variables and, in addition, is based on ranks as opposed to means. More detail on the processing and statistical analysis of data are provided in Chapter 3 and 4.

The quantitative analysis is primarily aimed at answering the first two research questions, and testing the two first hypotheses. Specific hypotheses were formulated for each of the operationalisations outlined in Table 1.1 (see Chapter 3 for more detail), in order to answer these two research questions. Qualitative analysis of data is used to answer the last research question, regarding reasons for the particular distribution of features in the three subcorpora.

1.7 Conclusion

Against the background presented in this chapter, Chapter 2 offers a more comprehensive theoretical overview of the concepts of translation expertise and the features of translated language within corpus-based translation studies. This discussion not only bridges the gap between the concepts of translation expertise and the features of translated language, but also provides the necessary background for the operationalisations of the various features selected as indicators of translation expertise. Chapter 3 is devoted to the methodology used in the study and, as such, provides more information on the corpus composition and the operationalisations selected for the study, along with specific hypotheses regarding the expected distribution for each operationalisation. In Chapter 4, the findings for the different operationalisations are presented and the results are discussed with reference to the hypotheses formulated for the various operationalisations. Chapter 5, finally, presents a summary of the results and also outlines the limitations of the study as well as further possibilities for corpus-based research on translation expertise.

CHAPTER 2: THEORETICAL FRAMEWORK

2.1 Introduction

The aim of this chapter is to provide context for and background to two sets of concepts that inform the theoretical background and methodology of this study. The first part of the chapter (see Section 2.2) focuses on the concept of translation expertise, while the second (see Section 2.3) investigates concepts surrounding the hypothesised features of translated language, as considered from a corpus-based approach. The purpose of the chapter is to explore the implications of a possible relationship between the two sets of concepts. Therefore, the discussion will focus on the theoretical and methodological implications of existing research on translation expertise for this study, leading to the identification of a current gap for product-oriented research in investigations of translation expertise.

2.2 Translation expertise

2.2.1 Introduction

Research on translation expertise has come into its own in the last decade (Göpferich, 2013:61), and numerous theorists have tried to conceptualise and understand the concept of translation expertise from various perspectives. Notwithstanding the attention the concept has received, it has proved difficult to circumscribe in a definitive way (Ulrych, 2005:17), and there is no single definition of translation expertise that is universally accepted within translation studies (Angelelli, 2009:13). As a result, the concept of translation expertise currently simply represents an assorted range of academic understandings about what has to be known, and consequently what has to be taught or learned, in order to become an effective translator (Shreve, 2002:154). Some scholars, such as Malmkjær (2008:293), have even argued that the concept appears to have reached a state of stagnation in that little advancement is currently made in the field.

A number of the contemporary definitions of translation expertise view translation as a specialised kind of communicative competence on the basis of a functional translation approach (Angelelli, 2009:25). Working within this approach, theorists such as Adab (2000), Orozco (2000) and Beeby (2000) consider translation expertise as the ability to interpret a text with a specific function in the source language and to adapt or transform this text into a translation that performs a particular function in the target language. Some of the other definitions of translation expertise refer to the

knowledge, skills, attitudes (Kelly, 2005:165; Munday, 2009:234), expertise and judgement (Fraser, 2000:53) that are declarative or procedural (Ulrych, 2005:18; Gile, 2009:9) and needed for professional work in the field of translation. Martín (2009:25) formulates a conceptualisation of translation expertise based on a cognitive-psychological viewpoint, arguing that the concept contains a number of features, namely the accumulation of heuristic rules which can be used to improve and simplify problem-solving approaches, the capabilities related to the performance of expert translators, such as domain knowledge, meta-cognition and meta-knowledge, and a collection of behaviours that enable economy in terms of skilled performance. In these types of approaches it is generally accepted that translation expertise "is a property in constant change throughout the lifespan of a translator" (Martín, 2009:26) which means that the concept should contain within it the quality of dynamism rather than stasis.

In this section, attention will be given to translation expertise from a conceptual and methodological perspective so as to provide the terminological and theoretical grounding required for this study. As a point of departure, three approaches to translation expertise will be outlined, namely the componential, cognitive and relevance-theoretical approaches. Moreover, there will be a focus on the methodological implications of existing studies of translation expertise, in order to establish the specific contribution of this study to the field.

2.2.2 Theoretical approaches to translation expertise

2.2.2.1 The componential approach to translation expertise

In an attempt to gain a better understanding of translation expertise, a number of scholars, including Neubert (2000), PACTE (2000, 2002, 2005, 2008, 2009, 2011), and Göpferich (2009, 2013), have proposed models of translation expertise. These models, in general, are based on the tenet that translation expertise consists of several different components, or subcompetences, located at different levels of knowledge and skills (Dimitrova, 2005:13), which are interrelated and interact with one another (Massey, 2005). These models are dynamic and open-ended in nature (Massey, 2005) and propose that the extent to which the different subcompetences are used will vary depending on the translation situation (Beeby, 2004:44).

Even though a number of different models have been proposed, a comparison of these models suggests a significant degree of overlap in the types of skills and knowledge generally regarded as the

components of translation expertise. For the purposes of the discussion here, these components are formulated as the linguistic, extra-linguistic, textual, psycho-physiological, instrumental, translation knowledge and strategic components (see Figure 2.1).

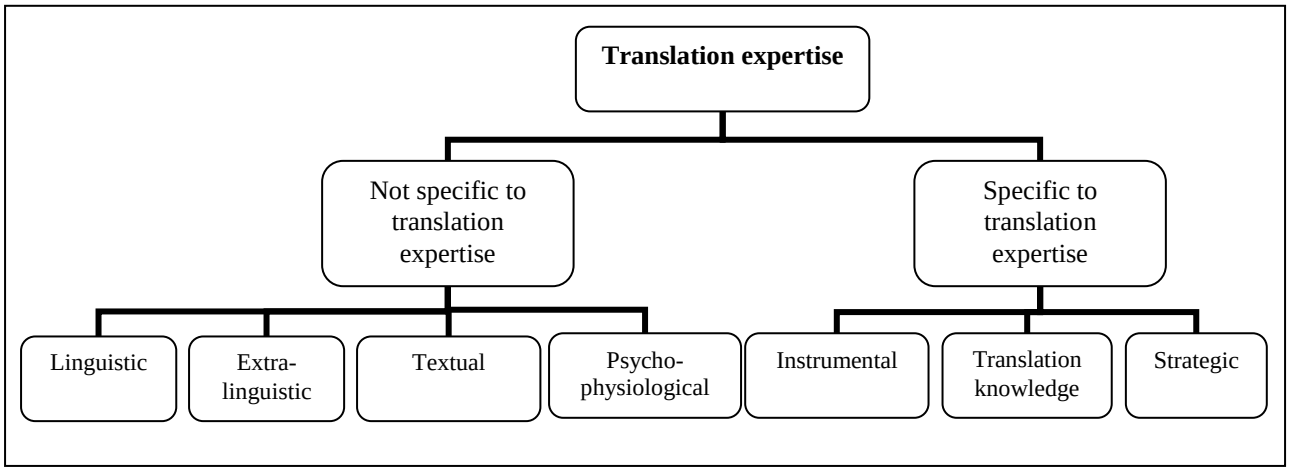

Figure 2.1: Diagrammatic representation of the componential approach to translation expertise

The first (and most obvious) of the suggested components of translation expertise is the **language or linguistic component** (Risku *et al*., 2010:91), specifically in terms of bilingual competence (PACTE, 2011:33), which is considered to be a vital, even undisputed aspect of translation expertise (Angelelli, 2009:31). This particular competence comprises skills that are invaluable to the overall quality of the translation product as it contains an element of textual and communicative competence in at least two languages and cultures (Göpferich, 2009:20), coupled with an awareness of discourse conventions in the involved cultures (Kelly, 2005:32). Moreover, lexico-grammatical understanding, expressed amongst others in graphemic, syntactic, morphological and vocabulary knowledge (Neubert, 2000:8) of the languages involved also form part of the language or linguistic component, together with knowledge related to register, style and genre conventions (Nord, 2005:161). A pragmatic element is also involved in the linguistic component (PACTE, 2009:208), which can be subdivided into illocutionary and socio-linguistic constituents. The former is knowledge of how language can be used to perform particular functions, such as to complain or apologise, whereas the latter has to do with knowledge of linguistic variation, such as dialects and regionalisms, and an awareness of figures of speech and cultural reference (Angelelli, 2009:24).

Translators who demonstrate expertise also have an **extra-linguistic competence** (Risku *et al*., 2010:91), which is based on the general and domain-specific knowledge needed to grasp the meaning

of the source text and to formulate the translation (Göpferich, 2009:21). Cultural and intercultural knowledge is considered to be an example of this competence (PACTE, 2011:33), as it is held that translators have to take into consideration the relevant beliefs and outlooks of the cultures involved (Neubert, 2000:10) to produce quality translations. In terms of translation expertise, the concept of culture not only refers to encyclopaedic knowledge of the geography, history or institutions of the involved cultures, but also to an awareness of the values, beliefs, perceptions, myths and behaviours that form part of these cultures (Kelly, 2005:32) and an understanding of their social environments and political situations (Nord, 2005:161). In addition to cultural knowledge, extra-linguistic competence also consists of subject-area knowledge (Gile, 2009:10), which involves an understanding of the source-text subject (Neubert, 2000:8). In this case, the subject knowledge may be encyclopaedic in nature and comprise specialist knowledge, although it is not restricted to the latter as "translators of seemingly non-specialist literary and related works" need it just as much (Neubert, 2000:8).

In addition, theorists such as Neubert (2000:8) and Angelelli (2009:33) argue that translation expertise involves a **textual skill**, which entails "the ability to string ideas together as a text" (Angelelli, 2009:32). This skill, in other words, is related to an awareness of internal structure of the source text and the effects that the structure has on meaning conveyed by the source-text author, which encourages translators to use their knowledge of similar target-language devices to render a comparable meaning in the target text (Angelelli, 2009:33). To some degree, this skill is influenced by the nature of translation as a system (Neubert, 2000:8). Translation is considered to be a system in that it is influenced by the norms, rules or conventions of text worlds which pertain to source and target languages and cultures (see also Toury, 2012). Thus, because of this nature of translation, the ability of translators to understand the normative usage of the languages involved (Neubert, 2000:8; Angelelli, 2009:33) is considered to be an important aspect of translation expertise. In this sense, knowledge of normative usage is also related to the subcompetence described as knowledge of translation, discussed below.

Psycho-physiological ability is considered to be another of the constituents of translation expertise, involving both cognitive and attitudinal components (PACTE, 2009:209). The cognitive aspect is related to factors such as attention span and memory, and involves abilities such as logical reasoning and creativity (PACTE, 2008:107). The attitudinal element is influenced by translators' perceptions

of themselves, on the basis of self-concept and self-confidence (Kelly, 2005:33), and also contains elements such as perseverance or determination and a critical attitude (PACTE, 2009:209). The way translators perceive themselves might contribute to their professional competence with respect to their negotiation skills (Kelly, 2005:33) and their ability to work as a team (Risku *et al*., 2010:89). Professional competence refers not only to knowledge about the work market (Munday, 2009:234), but also to interpersonal skills (Mackenzie, 2004:32) that are necessary to work effectively with other important role-players who are involved in the process of translation, such as revisers and terminologists, and other actors, such as clients and authors (Kelly, 2005:33).

The above-mentioned abilities and knowledge are considered to be vital to translation expertise, although PACTE (2005:611) argues that they are not necessarily specific to translation expertise. The skills and knowledge components they propose as defining features of the concept are the instrumental component, knowledge about translation, and the strategic component (PACTE, 2008:108). The **instrumental component** refers to the ability of translators to use communication technologies that are applicable to translation (Dimitrova, 2005:13), and is related to the skill of using reference and computer tools, documentary resources (Mackenzie, 2004:32), and databases and corpora (Göpferich, 2009:21) for professional practice while keeping the translation brief, if applicable, in mind (PACTE, 2009:215).

Knowledge about translation refers to awareness of the principles which serve to guide translation with respect to procedures, methods and processes (PACTE, 2011:33) and is influenced by an awareness of the norms that are applicable to the translation process. The concept of norms refers to the general ideas or values shared by a community that lead to performance instructions which specify what is applicable to and appropriate for specific situations, indicating what is allowed as well as what is permitted and tolerated in a particular behavioural dimension (Toury, 2012:63). The norms which influence translators' decisions form part of what is known as the basic initial norm and include an awareness that relevant similarity should be established between the source and target texts, that communication should be optimised and that the loyalty demands of the writer, the commissioner and the readership should be appropriately met (Chesterman, 1997:68-69). Preliminary norms are those that are related to translation policy, in terms of the texts chosen for translation in the first place, and whether direct or indirect translation (i.e. translation by means of a third, intermediary language) is tolerated, prohibited, permitted or preferred (Toury, 2012:82). Operational norms can be

either matricial in that they govern the existence of general translation material and the distribution thereof in the target text, or textual-linguistic in that the linguistic formulation of the translation is governed by them (Toury, 2012:60). These norms can be considered important in light of the knowledge about translation component as they influence the general strategies of translators and therefore may have an influence on the strategic component.

The **strategic component**, which Neubert (2000:10) refers to as the transfer component, is considered to be the most important constituent of translation expertise because it ensures the efficiency of the translation process in the face of translation as a problem-solving process (PACTE, 2009:209). Its role is to evaluate the translation problems it encounters and to suggest translation strategies which can be used to overcome these problems (PACTE, 2008:108). For this study, translation strategies are considered to refer to "the procedures (conscious or unconscious, verbal or non-verbal) used by the translator to solve problems that emerge when carrying out the translation process with a particular objective in mind" (Molina & Albir, 2002:508). These procedures, in other words, will activate other competences necessary to compensate for any shortcomings detected (Göpferich *et al*., 2011:60). During this process, the strategic component evaluates the efficacy of the translation process and the range of results obtained and tries to compensate for any shortcomings it identifies (PACTE, 2009:208). In addition to its problem-management role, one of the other central responsibilities of the strategic component is to focus on the types of translation expertise competences available to translators, on the basis of knowledge and abilities, and to select those which are most appropriate to the particular translation task or text segment which appears to be problematic (PACTE, 2011:33).

From the discussion of the componential models of translation expertise, it becomes evident that there are a number of different components located at different levels, in terms of knowledge or abilities, which are believed to contribute to translation expertise. An awareness of these abilities is of particular value to the field of translation training. Adab (2000:227), for instance, recognises the general usefulness of the componential models in light of its implications for translation training as it is argued that these models can be used to make students aware of the different components of translation expertise and how they relate to one another during the translation process.

Despite their evident usefulness, attempts at establishing componential models of translation expertise are not without problems. Beeby (2000:187) points out, for instance, that labelling the components of translation expertise does not enable one to know how the expertise is acquired or how it is used or practiced. This is complicated by the fact that many of the proposals lack empirical testing altogether, and specifically the validation that would be supplied by comprehensive and controlled empirical-experimental methods (Albir & Alves, 2009:64). While projects on translation expertise by PACTE (2000, 2002, 2005, 2008, 2009, 2011) and Transcomp (Göpferich, 2009) have moved in the direction of such empirical validation, they remain limited. Componential models have also been criticised because they separate some of the competences that some believe are not necessarily separate in reality, as a kind of heuristic tool. For instance, according to Martín (2010:155) it not clear why some competences such as linguistic and extra-linguistic knowledge are hypothesised to be different constructs "when current cognitive science assumes that meaning is encyclopaedic". The nature of the strategic competence is also questioned as it is not evident whether it is conscious or unconscious or which observable parameters correlate with it (Martín, 2010:155). Thus, it can be seen from this discussion that the main criticisms of the componential approach to translation expertise are the lack of empirical testing, together with questions regarding the separation of potentially interrelated competencies.

2.2.2.2 Approaches to translation expertise drawn from cognitive psychology

In addition to the componential models, some translation theorists have opted to explain the notion of translation expertise by means of cognitive psychology (Martín, 2010:155) and have argued that in order to gain an understanding of and to study the concept one has to be aware of the relationship between procedural and declarative knowledge (Martín, 2009:26). According to Robinson (2007:84) translation is characterised by two different processes and mental states, namely a subliminal flow state and a highly conscious analytical state, which are respectively analogous with the concepts of procedural and declarative memory or knowledge (Alves, 2005). Both of these types of knowledge are situated in long-term memory (Mondahl & Jensen, 1996:99) and the principal difference between them is how material is stored cognitively (Robinson, 2007:85), even though it is generally assumed that procedural knowledge develops out of an increase of declarative knowledge (Shreve, 2002:163). In the subliminal procedural state material is transformed into habit or second nature and this type of knowledge occurs in unanalysed knowledge structures, whereas in the analytical declarative state it is brought back out of habit and made subject to conscious analysis (Robinson, 2007:85). In the

procedural state, translators are able to use automatic information processing (Mondahl & Jensen, 1996:100) as they work mainly on intuition or "automatic pilot" (Chesterman, 2000:79), which means that they are able to work faster (Robinson, 2007:85). The declarative state, in turn, is based on controlled information processing (Mondahl & Jensen, 1996:99) and is much slower (Robinson, 2007:84) as translators have to draw on critical rationality (Chesterman, 2000:79). This relationship between the two different knowledge systems may be represented schematically as in Figure 2.2:

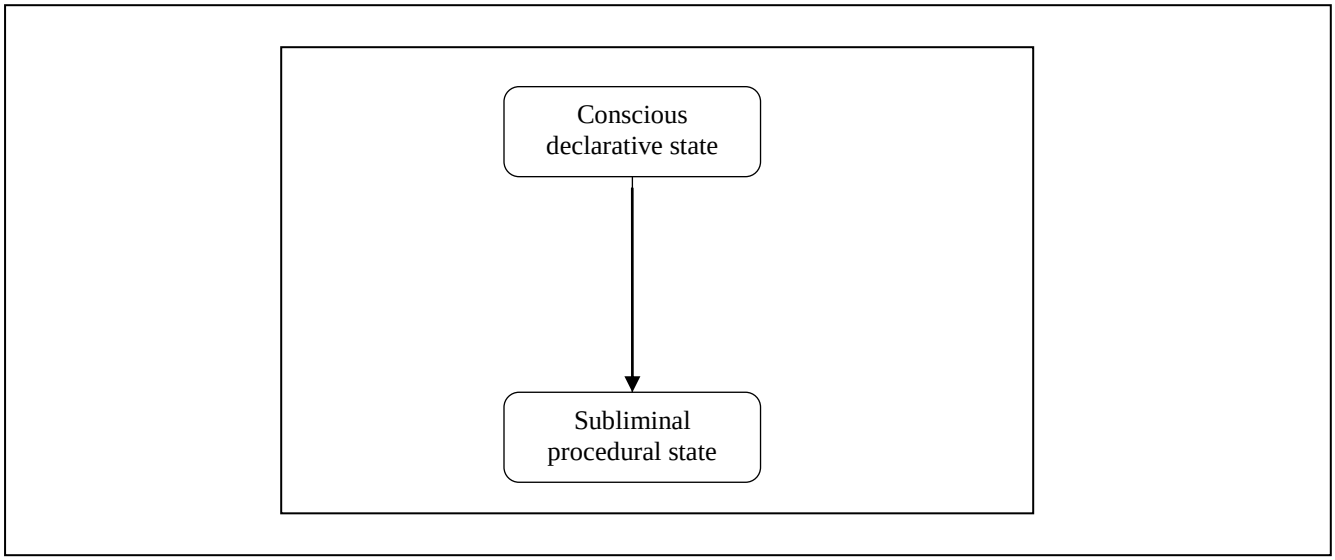

Figure 2.2: Knowledge systems that form part of translation expertise

Even though there is a general consensus that procedural and declarative knowledge are important factors of translation expertise, it appears that there is some discord regarding the extent to which of the two memory or knowledge systems are used by experienced and inexperienced translators. According to some theorists, experienced translators are expected to mainly use a procedural type of knowledge (Robinson, 2007:95; PACTE, 2011:33) and to switch to the more demanding and conscious mental processing of declarative knowledge only when the need arises (Martín, 2000:134), such as when they encounter translation problems or have to justify the solutions they have suggested in order to deal with a problem (Chesterman, 2000:79). According to this viewpoint, the processing of inexperienced translators will gravitate more towards controlled information processing in terms of declarative knowledge. Their performance will be characterised by effortful, conscious and deliberate processing, since novices have yet to acquire the procedural knowledge that develops with experience (Shreve, 2002:163).

However, the idea that experienced translators mainly work in an automatic, proceduralised fashion whereas inexperienced translators use more conscious processes and have a non-automatic approach to translation has been questioned. Jääskeläinen (2010:218) points out, for instance, that it has been found in previous studies that more successful professional translators tend to engage in more demanding, conscious processing activities, which suggests a reliance on declarative memory systems. These conscious processing activities result in what Tirkkonen-Condit (2005:407) refers to as a monitor model, Alves (2005) calls meta-reflection and Angelone (2010:18) labels meta-cognition. All three of these terms refer to the meta-cognitive ability of translators to critically monitor their own performance. It is held that due to the importance of this meta-cognition, experienced translators need "an extraordinary amount of declarative (explicit) knowledge; a type of knowledge that is conscious and deliberate" (Alves, 2005). In this view, inexperienced translators rely more on automatic processes, which means that they make few conscious decisions, leading to more automatic work (Alves & Gonçalves, 2007:49).

Although there is evidently discord about whether the cognitive processes of experienced and inexperienced translators are predominantly procedural or declarative, it is possible to solve the disagreement by looking at the argument from another perspective, in terms of a concept known as task-dependent flexibility (Jääskeläinen, 2002:111). According to Jääskeläinen (1996:67) translators are faced with translation tasks which are either typical and routine or novel and non-routine. With routine tasks, translators have the ability to work effortlessly and quickly (Jääskeläinen, 2002:111) as they have encountered similar texts on a regular basis and are familiar with their general characteristics and features, which has contributed to an accumulation of procedural knowledge. With non-routine tasks, however, translators are unfamiliar with types of texts and their associated features, and will therefore have to use a more conscious, effortful form of processing (Jääskeläinen, 1996:67) based on declarative knowledge. This results in translators shuttling between two different processes and mental states (Robinson, 2007:85).

Experienced translators have the advantage since they know when to use the procedural, routine approach or the declarative, non-routine approach as most effective for the task at hand (Göpferich *et al*., 2011:77), which implies that they choose between the two. Inexperienced translators do not have this task-dependent flexibility available to use (Jääskeläinen, 1996:67). It is important in this case to realise that "applying a routine approach to a non-routine task, or vice versa, may have undesirable

consequences" (Jääskeläinen, 1996:67). Therefore, it might be postulated that rather than looking at which type of knowledge is predominantly used by experienced and inexperienced translators to explain the notion of translation expertise, the focus should shift to acknowledge the fact that experienced translators may not only choose between the two but also decide which is best suited for the particular translation task at hand. This view can be linked to the strategic component of translation expertise discussed in Section 2.2.2.1, as it acknowledges that experienced translators know how to effectively draw on different cognitive strategies.

Schematically the task-dependent flexibility of experienced translators can be depicted as in Figure 2.3:

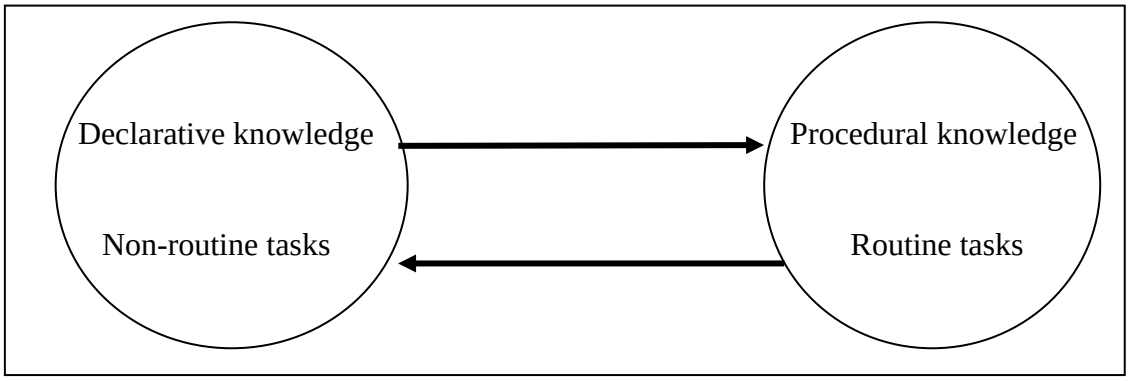

Figure 2.3: The task-dependent flexibility of translation expertise

The cognitive-psychology approach to translation expertise is not without its problems, however. As can be gathered from the discussion, one of the problems is caused by the fact that translation theorists cannot seem to agree on the extent to which the two memory systems are used by inexperienced and experienced translators. In addition to this, Kiraly (2000:2) objects to the fact that this approach to translation expertise is "epistemologically incompatible with a social process perspective". This is because the cognitive-science approach, on the one hand, is based on the assumption that knowledge and meaning are products of individual minds and are transferable, replicable, essentially static and independent of social interaction (Kiraly, 2000:2). The social-process perspective, on the other hand, assumes that knowledge and meaning are dynamic, intersubjective processes (Kiraly, 2000:2). Therefore, one of the main criticisms directed at the cognitive approach is that the assumptions on which it is based do not adequately consider the social processes involved in developing and maintaining translation expertise. This shortcoming is addressed in relevance-theoretical approaches to translation expertise.

2.2.2.3 Translation expertise from the perspective of relevance theory

Relevance theory is another cognitive approach that has been used to explain translation expertise. In the context of translation expertise, relevance theory postulates that translators do not just focus on external contextual factors or the representation of states of affairs, but rather have an awareness of different cognitive environments applicable to translation (Gutt, 2004:89). The concept of cognitive environments refers to the sum total of information available to people at a given point in time (Gutt, 2004:80). The concept of a mutual cognitive environment thus refers to information which communicators are expected to have in common with their audience (Gutt, 2004:81). From this perspective, translation is complicated in that translators have to be able to meta-represent the "mutual cognitive environment shared between the original communicator and original audience" while also being aware of the cognitive environments of their target audience (Gutt, 2004:82). In other words, within the relevance-theoretical paradigm, translation expertise is seen as dependent on the notion of meta-representational skills (Gutt, 2004:89).

According to Alves and Gonçalves (2003:8) "besides the mastery of many other cognitive skills, translators have to learn to manipulate more consciously conceptually and procedurally encoded information" along with contextually encoded information, if they are to meet the needs of the target audience. **Conceptually encoded information** expresses conceptual meaning. It "can be enriched and contributes to the inferential processing of an utterance" (Alves & Gonçalves, 2003:8). Conceptual information is encoded by open lexical items, such as verbs, nouns and adjectives **Procedurally encoded information** decisively contributes to the cognitive processing of an utterance by inflicting inferential constraints on the utterance (Alves & Gonçalves, 2003:8). In a linguistic sense, procedural information is mainly encoded by means of non-open linguistic items and structures such as tenses, determiners, word order and negation, amongst others. These procedural effects guide the reader towards the intended **contextually encoded information**. Context in relevance theory refers to the subset of a person's "assumptions about the world" and is used to interpret the message of the text (Malmkjær, 1992:26).

Based on the manipulation of the different types of encoded information, Alves and Gonçalves (2007:51) distinguish between the cognitive profiles of what they call broadband and narrow-band translators. Broadband translators are usually experienced translators who are able to effectively integrate the different types of encoded information into a coherent whole and their problem solving

tends to be driven by contextually embedded information. They are hypothesised to work mainly on the basis of communicative cues supplied by the source text, in other words, the contextualised meaning of words rather than the dictionary meaning. These communicative cues are then reinforced by the contextual assumptions derived from the translators' cognitive environments (Alves & Gonçalves, 2007:51).

Narrow-band translators, conversely, are mostly inexperienced translators who fail to connect the gap between procedurally, conceptually and contextually encoded information (Alves & Gonçalves, 2007:50). A study by Alves and Gonçalves (2003:15) demonstrates that narrow-band translators struggle to recognise procedurally encoded information because they lack the ability to effectively retrieve the communicative cues that are conveyed in the source text. In addition, they sometimes also extensively draw on their cognitive environments so as to process conceptually encoded information (Alves and Gonçalves, 2003:15). The differences between broadband and narrow-band translators are schematically represented in Figure 2.4. Focusing on the cognitive profiles of broadband and narrow-band translators is useful for the discussion of translation expertise since this approach allows one to concentrate on the performance traits that are characteristic of experienced and inexperienced translators (Alves, 2005).

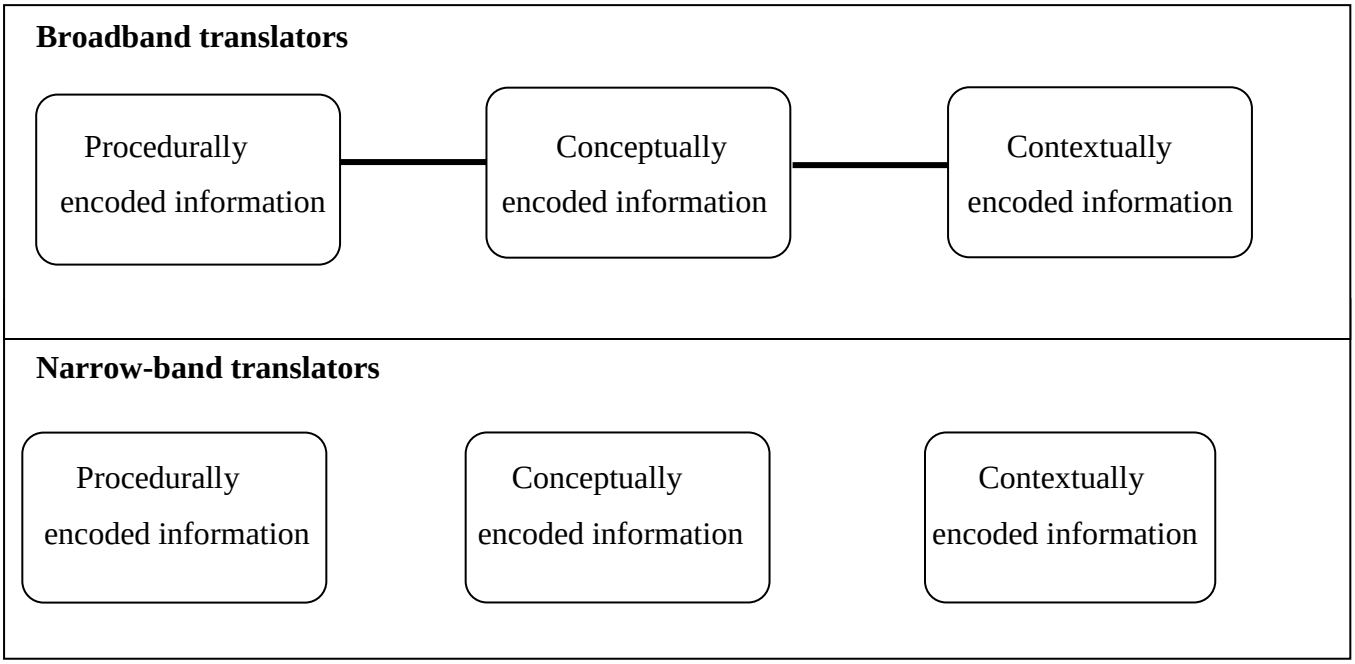

Figure 2.4: Differences in information encoding for broadband and narrow-band translators

However, despite the explanatory power of relevance theory in the context of translation expertise, the approach is not without criticism. According to Alves and Gonçalves (2010:283) relevance theory has been criticised by theorists who feel its applicability is too vague. In addition, relevance theory has been criticised for being largely producer-oriented, which means that, overall, less attention is paid to the receiver or the translation audience and more to the translator (Tanskanen, 2006:23). In terms of translation expertise this may hold the implication that less attention is paid to how translations produced by experienced and inexperienced translators are received and interpreted by their target audiences.

2.2.3 Implications of the three approaches to translation expertise

The approaches to translation expertise discussed above have several implications for research in the field. In the first instance, it should be noted that the different approaches to translation expertise have developed largely in isolation from one another, and their principles have not been combined or synthesised to a significant extent. In other words, even though studies based on these approaches to translation expertise are similar in the sense that they focus on the same field of enquiry, they have not been linked to one another in terms of the basic assumptions on which they are based. However, there is clearly some degree of overlap between the three approaches, which permits other studies in the same field to build on their general principles, using them as a point of departure to extend the overall understanding of translation expertise. This study will therefore refer to concepts related to cognitive psychology, use ideas about subcompetences involved in translation expertise, or borrow the fundamentals of relevance theory so as to account for some of the differences between experienced and inexperienced translators. As a number of studies have already yielded insights into the differences between experienced and inexperienced translators, some of these findings will be summarised and discussed in the next section in order to aid an understanding of the influence that experience has on the translation process and the translation product.

2.2.4 Differences between experienced and inexperienced translators

Generally, experienced translators have a holistic approach to translation which means that they are able to focus on the text to be translated as an organic whole (Angelone, 2010:34) and as such take note of the more complex or deeper structures of the source text since they see the translation task on an abstract level (Shreve, 2002:166). This holistic approach allows experienced translators to recognise problematic features of the text and enables them to use abstract methods to solve

translation problems and produce more effective target texts, such as reordering sentences in paragraphs or moving semantic units from one paragraph to another (Shreve, 2002:166). Additionally, they are able to formulate globally oriented strategies (Dimitrova, 2005:14), or strategies which emphasise the needs of the target audience with respect to general translation principles (Jääskeläinen, 1993:115). In this sense they can focus on the task problem areas which will have the greatest impact on the outcome of the target text and, consequently, they spend their decision-making efforts strategically and do not waste it on irrelevant detail (Tirkkonen-Condit, 2005:407).

In addition, it is presumed that the cognitive processing of experienced translators proceeds in an iterative and non-linear manner (Göpferich & Jääskeläinen, 2009:174), in that alternatives for a particular text segment are sought even when a translation has already been found (Séguinot, 2000:146). According to Séguinot (2000:146) this feature of the cognitive processes of experienced translators can be taken to indicate parallel processing in the sense that they are able work on more than one item at a time. As a consequence of such parallel processing, experienced translators produce tentative translation equivalents and also critically monitor these tentative solutions because they are more aware of translation problems (Göpferich & Jääskeläinen, 2009:174) and because of the development of meta-cognitive abilities (Alves, 2005). The meta-cognitive abilities encourage experienced translators to constantly check their target-language text output, regardless of whether any problems were involved (Lörscher, 2009:605).

With increased experience the text segments focused on, in terms of phrases, clauses or sentences (Lörscher, 1996:30), become larger (Göpferich & Jääskeläinen, 2009:174). This feature of cognitive processing allows experienced translators to adopt a sense-oriented approach to translation (Lörscher, 2009:15) which means that they can match the "sense of a source text translation unit to a unit with an equivalent sense in the target message" (Shreve, 2002:166), which allows for a more natural-sounding target text. Experienced translators can glean a large amount of knowledge from the source text and, consequently, focus on the source text itself and maximise the quantity of information to be inferred from the inter-textual, pragmatic and semantic dimensions (Tirkkonen-Condit, 2005:406) in combination with their capability to exploit their personal cultural and world knowledge (Jääskeläinen, 2010:221).

In contrast, inexperienced translators focus on the micro-level of the text (Snel Trampus, 2002:44) and are therefore not yet able to process it as an organic whole. Their focus and their translation process are primarily restricted to the level of individual words (Cintrão, 2011:89) as the succession of signs, in terms of sentences, clauses and words (Lörscher, 2005:13), creates a certain kind of irresistible pressure throughout the translation phase (Lörscher, 2009:13), resulting in the visibility of the source text structures in the translation (Shreve, 2002:166). Inexperienced translators' reliance on the source text is also thought to be determined by their general view of translation in that they generally consider translation to be a replacement of signs between two different languages (Cintrão, 2011:89), which means that the source-language text is credited a dominant role (Lörscher, 2009:13). Inexperienced translators therefore demonstrate form-oriented approaches to translation, as opposed to sense-oriented ones, as they tend to produce translations based on an exchange of language signs (Lörscher, 2005:605). Consequently, they use world and textual knowledge to a lesser extent when compared to experienced translators (Dimitrova, 2005:15).

As a result of their micro-level approach, inexperienced translators tend to focus on the superficial aspects of the translation problem (Shreve, 2002:166). They tend to rely on local strategies (Tirkkonen-Condit, 2005:407), or lower-level problem-solving activities or decisions which occur in relation to particular lexical items (Jääskeläinen, 1993:115), even though they sometimes end up with nonsensical translations (Jääskeläinen, 2010:211). Moreover, they problematise little and check the solutions to their problems concerning lexical equivalence and syntactic correctness to a lesser extent than experienced translators do (Jääskeläinen, 1996:31). Nevertheless, their reliance on the lexical level of the text does not mean that they do not engage in meta-cognition at all, because according to the hypothesis of Angelone (2010:19) both inexperienced and experienced translators have meta-cognitive ability. However, the difference is situated in the contour and the degree of meta-cognition used in their problem solving.

Inexperienced translators also have poor time allocation with respect to orientation, drafting (Alves, 2005) and editing or revision phases (Martín, 2009:30) and differ from experienced translators in the sense that they are more inclined to take risks (Künzli, 2004:45) even though they have a lack of self-confidence in terms of their problem-solving strategies (Angelone, 2010:32). This risk-taking tendency can be attributed to the fact that they do not face possible loss of credibility if their

interpretation is wrong and also because they have not gained enough knowledge to know where risk-taking is necessary and where it should be avoided (Künzli, 2004:45).

In general, studies such as those discussed in this section are based on two methodological approaches, namely process- and product-oriented approaches. These two methodological approaches will be examined in more detail in the next section along with their implications.

2.2.5 Methodological implications of existing studies

Current empirical research on translation expertise and its development focuses both on translation processes and translation products. The two foci may be seen as related to one another in that the observed properties of translations may be correlated with behaviour that is directly or indirectly observable throughout the translation process (Alves *et al*., 2010:110). In order to gain an understanding of the act or the process of translation itself, process-oriented methodologies are most obviously useful (Holmes, 1972:185). These methodologies are, in general, based on the principles of psycholinguistics or cognitive psychology (Halverson, 2008:212), and have the advantage that they allow direct access to the cognitive processing that occurs during translation. They therefore produce abundant process information (Alves *et al*., 2010:111). Some of the main instruments include think-aloud protocols (TAPs) and retrospection, interviews, dialogue protocols, eye-tracking, neuro-imaging techniques, questionnaires, commercial software programmes which record and monitor the activities of users (e.g. keystroke logging and screen recording), and direct observation charts that record the activities such programmes are not able to record (Albir & Alves, 2009:70; Göpferich & Jääskeläinen, 2009:171-174).

Studies based on process-oriented methodologies have made significant contributions to the field of translation expertise (Jääskeläinen, 2008:291; O'Brien, 2013:5). Some of these include the studies of PACTE (2003, 2005, 2009) from a componential perspective, the study of Jääskeläinen (1999) from a cognitive perspective and the study of Alves and Gonçalves (2003) from the viewpoint of relevance theory. However, it should be noted that process-oriented methodologies also have their drawbacks. Most of the abovementioned methodologies are problematic since they may unnerve translators who struggle with the artificiality of the test conditions which, in turn, might have an effect on the results (Angelelli, 2009:21). In addition, it has been noted that process-oriented methodologies based on verbal reporting are limited by the inevitable incompleteness of data, since they cannot provide more

than an "incomplete account of the processing involved in any cognitive task" (Jääskeläinen, 2008:291). Moreover, according to Olohan (2002:5), some process-oriented methodologies (such as TAPs) have also been criticised for their inability to give access to mental processes that are automated or subconscious.

In order to make up for some of the deficiencies of these methodologies, triangulation has recently been used as a methodological approach. Triangulation refers to "the use of two or more data acquisition methodologies within a single study to improve the quality, validity, and reliability of research findings" (Shreve & Angelone, 2010:6). In the context of research on translation processes, different methods offer different types of information that may illuminate different aspects of the translation process (Göpferich & Jääskeläinen, 2009:171). With these types of studies it is possible to triangulate data collected by process methods, as in Alves and Gonçalves (2003), who triangulate TAPs with process data from Translog, a computer program that records all keyboard activity. It is also possible to triangulate data collected by means of both process and product methods, as in Carl (2009), who triangulates both the process and product data supplied by Translog output.

Product-oriented methodologies, which are concerned with translations as linguistic artefacts (Halverson, 2008:212), have also been used to study translation expertise, albeit to a lesser extent than their process-oriented counterparts. Some examples of product-oriented methodologies include parallel and comparable corpora, translation quality assessment and product-oriented error analysis (Garant, 2008:33). However, these methodologies have their own set of problems. Alves *et al*. (2010:11), for instance, point out that one of the main disadvantages of this approach is that it constitutes post-hoc information: the product is the end result of the process that created it. This means, in other words, that processes have to be inferred from products, which is clearly not a straightforward matter. Temple and Young (2004:170) argue that the use of this type of methodology means that the translators who produced the texts do not generally have an active role to play in studies, as translation scholars tend to focus mainly on the translations themselves in terms of their relationships of equivalence. In this sense, the perspectives of the translators are lost, since there is no reflexive debate with translators about the choices they made when they produced the texts (Temple & Young, 2004:170). Nevertheless, despite these objections to product-oriented methodologies, they have the advantage that they provide the researcher with authentic data that were not influenced by the artificiality of test conditions.

One of the most popular product-based methodologies in current empirical research on translation is the corpus-based approach, which, according to Kenny (2008:62), has proved to be "a dynamic force in Translation Studies". Recently, there has been the development of sign-language corpora (Leeson *et al*., 2006), multimedia corpora of dubbed and original films (Valentini, 2006) and subtitle corpora (Armstrong *et al*., 2006), amongst others. However, much of the early work in corpus-based translation studies focussed on the identifiable, specific features that might be associated with the nature of the translation activity itself (Kenny, 2008:61; Delaere *et al*., 2012:137), which continues to be of interest to translation scholars, using increasingly sophisticated methods of analysis (see Delaere *et al*., 2012; Williams, 2005; Becher, 2011; Yuan & Gao, 2008). In the next section, the corpus-based approach to the study of the features of translated language will be discussed in more detail so as to gain an understanding of its applicability to the research questions informing this study.

2.3 The corpus-based approach and the features of translated language

2.3.1 Introduction: corpora and descriptive translation studies

Corpus-based translation studies as a field of research is informed by a research method in linguistics known as corpus linguistics, which entails the study of large computerised corpora of texts with computer programs (Kruger, 2002:70). A corpus refers to a text collection – which may consist of texts from a number of sources on a multitude of topics, by various writers and speakers – which is held in machine-readable format and can be analysed semi-automatically or automatically in a number ways (Baker, 1995:225). The corpus-based approach has recently been seen as representing a fruitful area of progress in translation studies (Baer & Bystrova-McIntyre, 2009:161), primarily because a corpus can offer large amounts of authentic product information with a high degree of ecological validity (Alves *et al*., 2010:111) – one of the reasons why this methodological approach was chosen for this study.

A primary concern with the use of corpora in the investigation of various aspects of translation is that a methodological gap frequently exists between the hypothetical construct under investigation and the linguistic phenomena extracted from the corpus to support the hypothesis (Alves *et al*., 2010:111). In addition, corpora have to be representative of a specific kind of language production and reception, which is problematic as the statistical concept of representativeness is "extremely difficult to apply to textual data, and [as a result] many commentators now prefer to aim for a 'balanced' sample of the language in which they are interested" (Kenny, 2008:59). In order to arrive at a more balanced sample

it may therefore be necessary to include both written and transcribed spoken language and within these a variety of text types which date from specific time periods (Kenny, 2008:59).

Despite some of the methodological limitations associated with the use of corpora to investigate translation phenomena, the emergence of corpus-based translation studies is generally regarded as a productive area of advancement in translation studies overall in that it enables scholars to pursue a range of agendas (Kenny, 2008:59). For instance, corpora have been used to uncover translation shifts and translation norms and to provide empirical information for machine translation systems (Kenny, 2008:61). Corpora have been instrumental in studies of the "universals" or features of translated language (see Section 2.3.2). They are frequently used as a source of information in contrastive linguistics, and may also contribute to the training and education of learner translators to help them find solutions to problems that are characteristic of translation (Malmkjær, 1998:535-536; Kenny, 2008:61).

Depending on the purpose of the study, translation scholars have a choice of two different types of corpus composition, namely a parallel corpus or a comparable corpus – both of which have advantages and disadvantages. They may also combine the two types of corpora to solve the problems associated with using one or the other. A parallel corpus is composed of two (or more) subcorpora: a subcorpus of source-language texts and a subcorpus (or subcorpora) of the translations of these texts in another language (or languages) (Baker, 1995:230). Most parallel corpora are bilingual and consequently allow the translation scholar to study a specific language pair, even though this kind of corpus can potentially contain translations into a variety of languages of the same source-language texts (Kruger, 2002:86). In general, the most important contribution of parallel corpora to translation studies as a discipline is that they allow theorists to establish objectively how translators are able to overcome translation difficulties in practice (Baker, 1995:231). However, one of the biggest problems associated with parallel corpora is that their construction is quite laborious and except for a few language pairs, such as English-Chinese and English-French, they are a scarce resource (Zanettin, 2012:33).

A comparable corpus is made up of two text collections, or subcorpora, in the same language: one which is composed of translated texts and the other which consists of comparable non-translated, or original, texts (Laviosa, 2002:36), where comparability may have to be ensured across various

dimensions, such as text type and historical period. The texts contained in a comparable corpus have no direct translational relationship in that "they are not translations of each other" (Zanettin, 2012:11). The principle behind the comparable corpus approach is that a comparison of non-translations with a corpus of comparable translations will highlight the features of the latter, which are explainable by the effect of translation, specifically (Bernardini & Zanettin, 2004:53). However, as Bernardini and Zanettin (2004:53) point out, the principle on which the comparable approach is based is problematic in practice, as it may be difficult for comparability to be ensured between the subcorpora.

The advent of large electronic corpora, together with the development of software needed to extract data from such corpora, has had a significant influence on reinforcing the pivotal function of description in translation studies by means of the development of a coherent methodology and the acquisition of new knowledge and a better understanding of translational behaviour (Laviosa, 2008b:122). The corpus-based approach has been particularly useful in the field of descriptive translation studies, which strives to describe, explain and predict translational phenomena (Toury, 2012:1) by unveiling the relationship between the process, product and function of translation (Laviosa, 2003:47). Within the broader project of translation studies, research based on the descriptive paradigm has an important contribution to make, since carefully performed descriptive studies constitute one of the best ways to test, refute, modify and amend the theories in terms of which the research is carried out (Toury, 2012:1). Descriptive translation studies has a reciprocal relationship with translation theory in that its empirical nature allows translation theorists to test translation hypotheses so as to determine whether there is support for their claims (Laviosa, 2003:46).

In the following section, the concept of features of translated language (as one of the key hypotheses of descriptive translation studies) will be explored in more detail, with reference to existing studies in the field. The section will focus on explicitation, simplification, normalisation and levelling-out, while the concept of interference will be briefly discussed where relevant. The ultimate aim of this section is to lay the groundwork for the argument that these features of translated language may be taken as indicators of translation expertise as a whole.

2.3.2 Universals, laws and features: recurrent patterns of translated language

One of the concerns of descriptive translation studies that the corpus-based approach has allowed theorists to investigate in a more rigorous way is the claim that translation demonstrates distinctive linguistic patterns or regularities. This is because translated language conforms to its own structural entailments and presuppositions which differ from non-translated language and, as such, constitutes a "third code" (Frawley, 1984:253). Before the widespread availability of corpora, such hypotheses were mostly tested by means of individual and combined textual analyses, as in the studies of Blum-Kulka and Levenston (1984), Blum-Kulka (1986) and Øverås (1998). However, with the development of electronic corpora and corpus-linguistic techniques, later descriptive theorists have further developed hypotheses regarding recurrent linguistic patterns in translation which are considered to be similar or closely related to the notion of the third code (Yajun & Zaixin, 2008:27).

One particularly influential hypothesis has been Toury's formulation of the laws of translation, which refer to "theoretical formulations purporting to state the relations between all variables which have been found relevant to a particular domain" (Toury, 2012:295). The two translational laws Toury proposes are the law of growing standardisation and the law of interference (Toury, 2012:303). The law of interference can be taken to refer to the transference of phenomena that are related to the source-text make-up to the target text (Toury, 2012:275), whereas the law of growing standardisation refers to the replacement of source-text items with target-text items that are more typical of the target-language repertoire (Toury, 2012:268). For Toury (2012:264), it is important to stress the probabilistic nature of translation laws as he feels that one should not only be able to describe and offer feasible explanations along with viable predictions for the occurrence of certain translation phenomena, but also to explain the occasional prediction failure that may occur. Laws, in other words, allow one to "to explain away (seeming) exceptions to a law with the help of another law, operating on another level" (Toury, 2004:29).

Another related influential hypothesis that focuses on the occurrence of recurrent translation phenomena is the concept of universals of translation proposed by Baker (1996). These universals refer to linguistic features that occur typically in translated rather than non-translated texts and are believed to be independent of the particular language pairs concerned in the interpretation process (Chesterman, 2006:3). The four potential universals proposed by Baker (1996:176) include simplification, explicitation, normalisation or conservatism, and levelling-out. These universals, in

terms of descriptive translation studies, have a predictive power in that it is expected that these universals will "occur in instances of translation that we have not yet encountered" (Kenny, 2001:54). It is the predictive power of the universals of translation that creates the impression that they resemble Toury's laws of translation, but the difference is situated in the fact that laws are more probabilistic and nuanced than many of the articulations of universal features of translation (Kenny, 2001:54).

The notion of translation universals has been criticised precisely for not making allowance for the impact of variable translation norms on translational behaviour and products. This is problematic because it is recognised that norms are determined culturally and socially, and change over time, while the concept of universals is more inflexible as it is based on the belief that universal features occur irrespective of the period, genre, language, the translator or ideologies (Tymoczko, 1998:653; Kenny, 2001:52). In other words, the relation between norms and translation phenomena may be subject to historical variation (Venuti, 2000:800), a fact which is not accounted for within the paradigm of translation universals. As a result, some translation scholars have opted to use different terminology to name the recurrence of certain linguistic phenomena, including "hypotheses" (Blum-Kulka, 1986), "core patterns" (Laviosa, 1998b), "laws" (Toury, 2012), "features of translation" (Olohan, 2004) or "regularities of translations" (Zanettin, 2012) . As the present study is not primarily concerned with the debate surrounding universalism, the term "features of translated language" will be used rather than "universals of translation".

In additional criticism of the concept of "universals", Chesterman (2004a:43) argues that Baker failed to make an important distinction in her original conceptualisation of the recurrent features of translated language. In this regard, Chesterman (2004a) argues for the need to acknowledge the role of different kinds of corpus composition in relation to hypotheses about the features of translated language. He points out that it is important to realise that the two types of corpus composition available for use are concerned with two separate key aspects in translation research, namely the translation's equivalent relation with the source text and the translation's textual fit with comparable non-translated texts (Chesterman, 2004a:39). Based on this, Chesterman (2004a:39; 2010:41-42) makes a distinction between source and target universals. S-universals, or the "characteristics of the way in which translators process the source text", are considered to be identifiable by a parallel-corpus methodology whereas T-universals, which are "characteristic of the way translators use the target language", may be revealed by using a comparable corpus (Chesterman, 2004a:39).

Despite the usefulness of Chesterman's distinction it should be noted that the distinction between S- and T-universals are not necessarily clear-cut. Even though some features can clearly only form part of one or the other grouping, such as interference (Toury, 2012) which can only be seen by means of a parallel corpus, others can fall into either category, such as simplification (Laviosa, 1998a) which can be studied by means of parallel or comparable corpora. Not surprisingly, the difficulties in making clear-cut distinctions have formed the basis of some criticisms of the concepts. Mauranen (2008:34), for instance, feels that even though the distinction between S- and T-universals has helped to clarify the discussion of features of translated language, the actual empirical research done in this respect has most frequently focused on T-universals. T-universals, in other words, have been studied more extensively with large databases than the S-universals. This is despite the fact that parallel corpora are available to examine S-universals (Mauranen, 2008:34).

Despite some of the objections to or problematic areas in the field, the study of the features of translation is generally considered to be beneficial to translation studies because, as pointed out by Chesterman (2004b:12), investigating these phenomena helps translation scholars and translators to gain a better understanding of the nature of translation. In this respect, one of the main reasons for investigating these features of translation is because it has the potential of raising translators' awareness of the unconscious or conscious processing involved in translation (Chesterman, 2010:46). This observation is based on an important idea: that these features may be the consequence of either deliberate translation strategies or due to largely unconscious cognitive processing that forms part of the complex nature of the translation activity (Olohan, 2001:423). A greater awareness of the outcomes of these processes may help translators to become more conscious of strategies and decision-making processes that contribute to translation quality (Chesterman, 2010:46).

2.3.3 The features of translated language

2.3.3.1 Explicitation

(a) Overview

Explicitation as one of the hypothesised features of translated language was first introduced to translation studies by Vinay and Darbelnet (1958:342) who define it as a "a stylistic translation technique which consists of making explicit in the target language what remains implicit in the source language because it is apparent from either the context or the situation". Explicitation, then, is a phenomenon that leads to target texts stating the information of the source text in a more explicit

form when compared to the original and, as such, it is assumed that there is a correspondence between source text implicitness and target text explicitness (Saldanha, 2008:21). Implicitation is the inverse of explicitation and occurs when ambiguous target text items are used to replace explicit source text items (Øverås, 1998:562). One of the key observations in studies of explicitation is that explicitation from the source language to the translation is not necessarily counterbalanced by the occurrence of implicitation in the direction from the translation to the source text – a phenomenon which is known as the asymmetry hypothesis (Klaudy, 2008:107). This phenomenon is based on the principle that is possible to distinguish between obligatory explicitation – where the occurrence of explicitation in one direction always corresponds to the occurrence of implicitation in the other direction – and optional explicitation – where the correspondence is not constant or exact (Pym, 2005). The asymmetrical relationship is believed to occur because translators, if given a choice, will prefer operations which involve explicitation in their translations and will fail to use optional implicitations.

In addition to the view that explicitation is related to the notion of implicitation, it has also been theorised that explicitation can be related to the notions of redundancy, self-referentiality and retrievability. As far as the first-mentioned concept is concerned, it has been argued that explicitation results in a target text that is more redundant than the source text, which is known as the explicitation hypothesis as formulated by Blum-Kulka (1986). According to this theory, the process of translation results in an increase in the degree of cohesive explicitness from source-language to target-language texts regardless of differences between the stylistic preferences and grammatical systems of the two languages involved. In other words, according to the explicitation hypothesis one can distinguish between shifts that occur as a consequence of differences in explicitness conventions between the source and target language and translational explicitness which occur as a result of the process of translation. In terms of Blum-Kulka's (1986) hypothesis, only translation shifts that occur in excess of what is required by the communicative conventions and the target-language system can be considered to be legitimate instances of explicitation.

Even though it should be noted that Blum-Kulka's explicitation hypothesis has been subject to criticism from Séguinot (1988:108), who sees the definition as too narrow as explicitation does not necessarily involve redundancy, and by Pym (2005) and Kamenická (2007:46), who feel that Blum-Kulka formulated her key concepts rather loosely, Baumgarten *et al*. (2008:184) find it to be an

important hypothesis. This is because Blum-Kulka's hypothesis distinguishes between the factors that contribute to explicitation in terms of explicitness conventions and translational explicitness, a distinction which many translation scholars frequently do not acknowledge (Baumgarten *et al*., 2008:184). Becher (2010b:2), however, cautions that this distinction does not mean that explicitation only occurs either as language-pair specific or as translation-inherent. This is because it is important to realise that it is possible for language-pair specific explicitation and translation-inherent explicitation to be compatible with one another (Becher, 2010b:2).

Klaudy's (2008) approach to explicitation is similar in that she also recognises the need to distinguish between different types of explicitation, and acknowledges the occurrence of obligatory, optional, pragmatic and translation-inherent (optional) explicitation. In this scheme, the occurrence of obligatory explicitation is attributable to differences in the semantic and syntactic structures of the languages involved, and without explicitation, the target-language sentences would not be grammatical (Klaudy, 2008:106). In general, semantic explicitation consists of more specific words that are used in the target text, whereas an increase in the number of words will most likely be as a result of syntactic explicitation (Klaudy, 2008:106). The incidence of optional explicitation is believed to be dictated by the stylistic preferences applicable to different languages as well as the differences involved in text-building strategies (Klaudy, 2008:106). Optional explicitation is considered to be optional because grammatically correct target-language sentences can be constructed without its use, even though the text might come across as unnatural or even clumsy (Klaudy, 2008:106). Translation-inherent explicitation, meanwhile, is attributable to the translation process, while pragmatic explicitation is regulated by differences between cultures. In the latter case, some information may not be shared by the source and target cultures, which consequently means that translators have to add explanations or additional information to their translations (Klaudy, 2008:107).

In general, explicitation has frequently been equated with addition strategies (Klaudy, 2008:104), as with Klaudy's conceptualisation of pragmatic explicitation, for instance. As far as the relationship between explicitation and addition is concerned, some scholars have regarded explicitation as the more specific and addition as the more generic concept, such as Nida (1964:228), whereas others have seen explicitation as incorporating the concept of addition which means that explicitation can be regarded as the broader concept, such as Séguinot (1988:108). In general, these types of

conceptualisations have led to the view that explicitation will result in translations that are more informative than their source texts because there is an addition or an "introduction of extra information" (Olohan & Baker, 2000:142). However, Saldanha (2008:23) challenges the assumption that explicitation results in increased informativeness. The problem, according to Saldanha (2008:33), is that self-referentiality occurs when translators wish to clarify information by adding a gloss or contextual information. This gloss is believed not to introduce or add new information but merely to reword or rephrase the information already present since "the meaning of the words are [sic] inscribed in the word itself" (Saldanha, 2008:25). In this sense, it is necessary to acknowledge that it might be problematic to assume that the degree of informativeness will increase as a result of explicitation based on the notion of addition.

Kamenická (2007:51) is of the opinion that one can distinguish between the concepts of explicitation and addition by means of the notion of retrievability. This means that the occurrence of explicitation or addition depends on whether or not the information that forms part of the translation unit is retrievable from the source-text context. If the information is retrievable, explicitation has occurred (Becher, 2010a:2), whereas addition has occurred if the information is irretrievable from the source-text context (Kamenická, 2007:51). This view of addition can be extended to the notion of implicitation – which is often equated with omission – in that the incidence of implicitation or omission is dependent on whether the "information that marks the locus of the translation shift in the [source text] surface structure can or cannot be retrieved from the [target text] context respectively" (Kamenická, 2007:51).

In addition to the general conceptualisations of explicitation in terms of redundancy, self-referentiality and retrievability, it should be noted that explicitation can have different interpretations based on the corpus methodology used by translation scholars and whether they study explicitation as a translation process or product. In terms of the translation process, explicitation involves a shift that concerns the content or structure, and aims to resolve ambiguity or increase and improve cohesiveness by adding linguistic and extra-linguistic information because translators either consciously or unconsciously strive to meet expectations of the target readers (Pápai, 2004:145). Explicitation viewed from a product-oriented perspective presupposes a higher degree of explicitness in translations when compared to non-translated texts, which might manifest as linguistic features that are used at higher frequencies in translated texts than in non-translated texts, or as an addition of

linguistic and extra-linguistic information (Pápai, 2004:145). In this sense, studying explicitation as part of the translation process is most likely to be accomplished by a parallel-corpus approach whereas studying explicitation as part of a product-oriented approach will most likely involve the use of comparable corpora. Explicitation may therefore be conceived of as both an S-universal and a T-universal.

(b) Results of previous studies

In their studies of explicitation, some translation theorists have been aware of the importance of distinguishing between different kinds of explicitation shifts, particularly those that are caused by differences in language systems and those that are attributable to the translation process. The study of Øverås (1998) is an example of one investigation utilising a bidirectional parallel corpus that explicitly set out with this distinction in mind. In this investigation, Øverås (1998) uses Norwegian original novels with their respective translations into English, and English original novels with their translations into Norwegian to examine both explicitation and implicitation shifts, focusing particularly on the addition and specification of grammatical and lexical ties to investigate explication, and omission to study implicitation. In her investigation, addition is conceptualised as the insertions of conjunctions in the translation that are not present in the source text (Øverås, 1998:562). In terms of grammatical ties, addition refers to the insertion of target text conjunctions that are not present in the source text, whereas in terms of lexical ties, addition is seen as the replacement of a source text Ø-item with "a constituent or proposition" (Øverås, 1998:565). Specification, in the form of both grammatical and lexical ties, focuses on explicitation shifts which result from either substitution or expansion (Øverås, 1998:563). While these specific shifts are easily identified, it is more difficult to say whether they reflect translation-inherent explicitation or whether they are the consequence of differences in the language systems. In this study, lexical explicitation was found to occur to a greater degree than grammatical explicitation, while explicitation overall was more prevalent than implicitation in both translation directions.

The study of explicitation by means of a comparison of source and target texts, in other words with a parallel corpus, can yield interesting results. There have also been a number of attempts to study explicitation in translation by using comparable corpora. Mutesayire (2004) studies reformulation as a kind of explicitation, and therefore focuses on the identification of lexical items typically used to introduce reformulation in English, such as in other words, namely and that is to say. Her analysis

indicates that reformulation markers tend to be used significantly more in a subcorpus of translated fiction than in a subcorpus of comparable English non-translated fiction, which is taken to be indicative of a higher degree of explicitness in translated language. Likewise, Williams (2005), using a corpus of government texts with a bidirectional design of translated and non-translated English and French texts, finds evidence for explicitation (operationalised as syntactic element disambiguation) being more common in translated language, since optional syntactic elements appear more frequently in both the French and the English translated texts in her corpus. Optional syntactic elements refer to elements that can be omitted without necessitating a change to the basic syntactic structure of the sentence, and include that and which constructions (Williams, 2005:114).

Olohan and Baker (2000) also study explicitation in terms of optional syntactic elements by comparing English translations taken from the Translational English Corpus[5] with comparable literary and creative source texts selected from the imaginative section of the British National Corpus. In their investigation, they analyse the frequency of the optional *that* complementiser in both corpora for the various forms of the reporting verbs TELL and SAY. They find the *that* connective to be more frequently used in the translational subcorpus than in the non-translated subcorpus. That is also found to occur more often than zero for all of the verb forms in the non-translated subcorpus with the exception of *says* and *said*. Becher (2010b:5), however, finds these results ambiguous as a number of the source languages contained in the corpus, such as French or Czech, need a subordinator following reporting verbs. Consequently, their results might be interpreted as indicative of source-text interference as opposed to "evidence of inherent, subliminal processes of explicitation in translation" (Olohan & Baker, 2000:143).

Olohan's (2002) study is similar to the one conducted by Olohan and Baker (2000) in that she also focuses on optional syntactic elements. The Translational English Corpus is used along with a subset of imaginative texts taken from the British National Corpus to test the hypothesis that translated texts will manifest a higher frequency of optional elements because translations render the grammatical relations in a more explicit manner than original texts. In the study, a lower incidence of contractions in the English translated subcorpus was noted compared to the non-translated subcorpus, and the non-translated English texts were more likely to omit that-formulations.

[5] The Translational English Corpus (TEC) is a collection of contemporary written translations made into English from a variety of source languages. It is specifically designed for the study of translated texts. It is held at the Centre for Translation Studies at the University of Manchester Institute of Science and Technology (Olohan & Baker, 2000:151).

The study of Kruger and Van Rooy (2012), like the study of Olohan (2002), also focuses on the occurrence of the optional *that* complementiser and contractions in their corpus which consists of a subcorpus of texts translated from Afrikaans and European languages into English in South Africa as well as a reference subcorpus of texts selected from the International Corpus of English[6] for South Africa. Even though the optional *that* is included at roughly the same frequency in the translated and non-translated subcorpora, it is found that there is a statistically significant effect for corpus for the instances where the *that* complementiser is omitted. The reference subcorpus omitted *that* significantly more, meaning that the translation subcorpus is more explicit in comparison. It is particularly in the registers of persuasive writing and reportage, and to a lesser extent popular writing and business letters, where the translation subcorpus almost consistently avoids omitting *that* as a complementiser. Their results for contractions, however, do not reach statistical significance.

The study by Pápai (2004) is one of few that aims to investigate explicitation both as an S- and a T-universal, using both a parallel and a comparable corpus. Her study makes use of a subcorpus of non-translated English texts and their translations into Hungarian as well as a subcorpus of original Hungarian texts. The parallel corpus consists of the English texts and their Hungarian translations whereas the comparable corpus is made up of the Hungarian original texts that are compared with the Hungarian translations. The texts used range from literary to technical non-literary texts, and the study focused on the modification and addition of punctuation marks, and the addition of derivatives, conjunctions, conjunctions along with cataphoric reference, and discourse particles (Pápai, 2004). The operationalisations used are considered to be suitable for both the parallel and comparable corpora. Her findings suggest that both the parallel and the comparable corpora displayed a strong tendency of translation-related explicitation, even though it was expected that the English into Hungarian parallel subcorpus would contain more implicitation shifts due to explicitness conventions in English (Pápai, 2004). As a result, it is concluded that the findings support explicitation as a recurrent feature of translated language (Pápai, 2004:159). Interestingly, her study did not find any clear-cut differences in the degree of explicitness between literary and non-literary texts (Pápai, 2004:160) even though it was hypothesised that the explicitness degree in scientific texts would be higher than that of literary texts (Pápai, 2004:145).

[6] The ICE (Corpus of International English) project was initiated in 1990, as an attempt to collect material for comparative studies of English across the world. Research teams are compiling one million-word corpora of spoken and written English of various national varieties, consisting of spoken and written texts produced since 1989 (see ICE, 2010 for more information).

Jiménez-Crespo's (2011) study is a response to Chesterman's (2004a:47) suggestion that the explicitation hypothesis should be tested on various translation types and, as such, tests the hypothesis against a comparable web corpus. The field of web localisation was chosen for three reasons. Firstly, it was postulated that explicitation ought to be present equally in both current and future translation types if it is a general or universal tendency. Secondly, explicitation may be expected to occur to a lesser degree in web translations, due to web usability guidelines and space constraints; and thirdly, studies based on comparable web corpora have produced evidence that contradicts other features of translation, such as normalisation. The Spanish Web Corpus was used in the study as a representative collection of texts of original Spanish, together with a subcorpus consisting of localised corporate websites of United States companies which are addressed to Spanish customers. The analysis consisted of two stages. In the first stage, the entire corpus was used to analyse syntactic explicitation in terms of optional Spanish personal pronouns whereas in the second stage, all navigation menu terminology was analysed. The study confirmed the explicitation hypothesis for optional Spanish syntactic items, namely articles and personal pronouns, that act as subjects for both stages of the analysis.

However, there are also studies which do not find evidence for explicitation. Puurtinen (2004) studied clause connectives in both translated and non-translated Finnish children's literature, and found no clear evidence for explicitation as a recurrent feature of translated language in terms of the translated subcorpus favouring connectives more than the non-translated subcorpus. Instead, some connectives were found to be more frequent in Finnish translations, others in Finnish non-translated texts and some to occur roughly equally in both the corpora used (Puurtinen, 2004:170). Nonetheless, it may be posited that it is possible that the results of the study were influenced by the nature of the literature studied in that the source texts might already have a high level of explicitness (which is favoured in texts for young children). This degree of explicitness might be such that it leaves no room for even more explicitation.

Becher (2010a:16) points out that there have been various methodological shortcomings associated with studies of explicitation. In his opinion, all the studies that focus on translation-inherent explicitation fail to make allowance for other possible interfering factors, such as source-language interference, different kinds of explicitation, or the effect of other potential features of translated language. In addition, many of the studies which offer a definition of explicitation do not adhere to it

while others do not even supply one, which carries the risk that pseudo-explicitations will be counted as examples of explicitation (Becher, 2010a:16-17; Becher, 2011b:28). In one of Becher's (2010b) studies, the explicitation hypothesis is tested by examining the German deictic adverb, *damit*, which lacks an equivalent expression in English. In order to proceed with the examination, a corpus is used that consists of English texts and their German translations as well as comparable non-translated German texts, with all the texts in the genre of popular science. Evidence is found for three of the four types of explicitation identified by Klaudy (2008), namely obligatory, optional and pragmatic explicitation, but translation-inherent explicitation proves to be problematic since it seems to be indistinguishable from optional explicitation. As a result, Becher (2010b:19) questions whether translation-inherent explicitation exists at all.

The problematic nature of translation-inherent explicitation leads Becher (2011) to question some of the possible motivations for translators to use explication – particularly to discover if its occurrence is attributable to the notion of translation-inherent explicitation. In this study, he uses a bidirectional parallel corpus containing business-text translations between English and German, and focuses on the addition of connectives. More explicitations are found to occur in the English to German translation direction and, in addition, it is established that explicitations are not necessarily counterbalanced by occurrences of implicitations, which provides support for the Asymmetry Hypothesis. Based on the findings, Becher (2011) argues that the majority of the shifts identified were the result of English-German contrasts in the sense of lexis, syntax and communicative norms, and argues that the idea of translation-inherent explicitation is therefore unfounded.

2.3.3.2 Simplification

(a) Overview

The feature of translation known as simplification was first defined by Blum-Kulka and Levenston (1984:119) as the "process and/or result of making do with less words", which was hypothesised to be influenced by the context and purpose of language use. According to Blum-Kulka and Levenston (1984:119), simplification is a feature of the language use of children who are acquiring their first language, of the writing and speech of second-language learners, of pidgins, and of texts prepared by native speakers for second-language learners and by speakers who are fluent in a particular language communicating with individuals who are less fluent in that language. In terms of translation, Blum-Kulka and Levenston (1984:119) believe that simplification occurs when a lack of exact equivalents

in the target language results in the use of similar devices as in other types of lexical simplification mentioned above, which therefore makes the communicative intent easier for the addressees to understand. This conceptualisation, therefore, may be taken to be more of an S-universal interpretation of simplification in that it centres on the idea that simplification is a consequence of the way in which the source text is processed by translators.

Baker's (1996:172) notion of simplification, although similar in nature, can be considered to be based more on a T-universal approach. She sees simplification as a technique that translators use to simplify the message and/or the language both lexically and syntactically for the target audience, to make "things easier for the reader" (Baker, 1996:182). She argues that this technique could result in shorter sentence lengths in translations (Baker, 1996:181) and lower type-token ratios and lexical density ratios in translations (Baker, 1996:183) – all of which have been used as operationalisations in corpus-based studies of simplification.

In addition to Baker's understanding of simplification as a T-universal, Laviosa has also approached simplification as a T-phenomenon that is discoverable by means of a comparable corpus. Laviosa (1998b:557) has postulated that certain "core patterns of lexical use" are indicative of the higher simplicity of translations when compared to non-translated text. In this regard, translations are expected to share four simplification characteristics: a lower degree of lexical density in terms of the ratio of lexical to grammatical words, in that more function words are used than lexical words; a relatively greater proportion of high-frequency words than low-frequency words; a relatively higher occurrence of the most frequent words; and less variety in the most frequently used words (Laviosa, 1998a:103). Translations, in other words, are expected to be typified by a relatively low information load and a limited range of vocabulary (Laviosa, 1998a:103), all of which contribute to simplification in translation.

In general, simplification is often thought of as a T-universal, and most theorists who have studied it have made use of comparable corpora. However, simplification might also be approached from an S-universal approach in that it has been posited that simplification is related to the concept of interference, and specifically to the Unique Items Hypothesis. According to this hypothesis, features of the target language that lack straightforward source-language linguistic counterparts (so-called "unique items") have a tendency to be under-represented in translation (Tirkkonen-Condit,

2002:207), leading to a lower number of "unique items" in translated texts than in comparable non-translated texts. This tendency may be seen as related to the law of interference (Kujamäki, 2004:188). According to this law, features or phenomena associated with the composition of the source text have a tendency to be transferred to translations either negatively, in that they occur as deviations from the normal practices of the target text system, or positively, in which case features are selected which do exist in the target language repertoire (Toury, 2012:303).

The law of interference and the Unique Items Hypothesis are related to simplification in the sense that translators tend to simplify by keeping closely to the make-up of the source text and consequently forget about the other alternatives available, which leads to them underutilising the resources of the target language according to its systemic possibilities, in comparison with authors of original texts (Eskola, 2004:96). According to Eskola (2004:96), it can be hypothesised that "translations tend to under-represent target language-specific, unique linguistic features and overrepresent features that have straightforward translation equivalents which are frequently used in the source language (functioning as some kind of stimuli in the source text)", which serves as evidence of the occurrence of a particular kind of simplification. Therefore it may be suggested that simplification can also be studied as an S-phenomenon as its occurrence might be attributable to the fact that the source text has an influential effect on the interpretation process of translators.

In addition to its relationship with interference, it has to be pointed out that simplification as a feature of translation has an ambivalent relationship with explicitation (Yajun & Zaixin, 2008:29), with some even considering it to be the obverse of explicitation (Williams, 2005:38). The reason for this ambivalent relationship is because, on the one hand, simplification contradicts the overall notion of explicitation whereas, on the other hand, they are related to one another. Pym (2008) argues that a relationship of dissimilarity exists in that simplification is hypothesised to shorten sentence lengths whereas explicitation in general results in longer sentences. Furthermore, it becomes even more problematic to recognise the line that divides the concepts since both of these translation features make texts easier to read and to understand (Pym, 2008), and there may even be a causative relationship between them. For example, simplification might occur as a result of explicitation strategies, as explicitation results in more lexical repetitions and, consequently, in less varied vocabulary (Pápai, 2004:159). More specifically, it may be stated that "all shifts [in terms of

explicitation] inevitably lead to lexical repetitions [and] consequently to simplification in the vocabulary" (Pápai, 2004:160).

In addition to its difficult relationship with explicitation, the simplification hypothesis is also problematic in the sense that its essence contradicts one of the proposed features by which it may be detected. As stated, simplification is based on the notion that a text is made easier for the target audience to process and understand. However, as Puurtinen (1998:525) points out, with reference to Finnish texts, this claim may well be unfounded. Shorter sentence lengths are often taken to be indicative of simplification; however, shortening sentences may require translators to make use of more complex, concise syntactic structures (Puurtinen, 1998:525). The use of these complex structures might make translations more difficult to understand, contradicting the idea that simplification equals a text that is easier to understand (Puurtinen, 1998:525). This argument may be extended to other languages, too – which could undermine one of the fundamental assumptions of simplification.

(b) Results of previous studies

A number of studies have aimed to uncover the extent to which simplification can be considered a feature of translated language, with some variability in findings. Most of these studies have used a comparable rather than parallel-corpus design to investigate simplification (Castagnoli, 2008:21) and therefore it can be deduced that simplification has mostly been studied as a prospective T-universal. Laviosa (1998a, 1998b), for instance, studied simplification by means of a subcorpus composed of literary and newspaper texts of translated English, and a comparable subcorpus of non-translated English, to test three hypotheses, namely that the translation corpus would be characterised by lower lexical variety in that the vocabulary range would be narrower (measured by type-token ratio); that it would demonstrate a lower information load in terms of "a lower ratio of lexical to running words" and that it would have shorter sentence lengths overall (Laviosa, 2002:61). The first two hypotheses were confirmed, whereas the last one was found to be true only of newspaper texts and not literary texts.

A study by Pastor *et al*. (2008), with Spanish translations done from American and British English and a comparable subcorpus of original writing in Spanish, has provided support for Laviosa's first hypothesis, supplying more evidence for the claim that the vocabulary of translated texts is less varied

in comparison with non-translated texts. Pastor *et al*. (2008) found additional evidence for the simplification hypothesis, observing that the translated texts in their corpus were characterised by shorter sentence lengths. However, they also noted that the translated texts in their corpus featured a "significantly smaller proportion of simple sentences" (Pastor *et al*., 2008), which means that a greater number of complex sentences were used. This finding can be linked to Puurtinen's (1998:525) view, briefly discussed above, that shortening sentences might result in more concise and complex syntactic structures that are more difficult to understand, which, in turn, undermines the essence of simplification as a feature of translated language.

Studies by Yuan and Gao (2008) and Williams (2005), however, do not provide consistent evidence for the occurrence of simplification. In Yuan and Gao's (2008) study, which used a subcorpus of Chinese fiction translated from German, Russian, French and English and a comparable subcorpus of original Chinese writing, it was found that sentence lengths and lexical density in the translation subcorpus were higher than those in the non-translated subcorpus (Yuan & Gao, 2008:22). They conclude their study by noting that the results appear to be influenced by the grammar and vocabulary of the specific language involved, and not so much by the non-translated or translated status of the subcorpora (Yuan & Gao, 2008:22). In Williams's (2005:208) study, findings were reversed in the two languages used in her study: the translated English texts had longer, more complex sentences and lower vocabulary ranges and information loads than comparable original English texts, whereas translated French texts had shorter, less complex sentences and larger information loads and vocabularies than comparable original French texts. These findings, therefore, lend more support to the idea that language-specific conventions may have an impact on how and the degree to which the features of translated language occur in translations.

Kruger and Van Rooy (2012) did find evidence for simplification as measured by type-token ratio, with their translated subcorpus having a significantly lower average type-token ratio than the non-translated reference subcorpus, which means that the non-translated subcorpus has a more varied vocabulary overall. However, the translated subcorpus had a higher average word length than the non-translated subcorpus, but the difference was so small that the main effect was not statistically significant. This study therefore also yields mixed findings for simplification as a feature of translated language.

2.3.3.3 Normalisation

(a) Overview

Systems theorists in translation studies have recognised the importance of acknowledging the fact that translation forms part of a literary system that is subject to the social, historical and literary dynamics of a specific culture (Even-Zohar, 1978:200) and, as such, is influenced by the norms of the target culture (see Section 2.2.2.1).

One of the ways in which these norms are thought to have an impact on translation is that a greater degree of conformity to convention is evident in translated texts when they are compared to original texts produced in the same language (Williams, 2005:17). This conformity results in a tendency for typical features and patterns of the target language to be exaggerated in translation. This kind of conventionalising tendency is most evident in terms of punctuation, grammatical structures (Baker, 1996:183), idiomaticity (Baker, 2007:14) and lexis (Kenny, 2001:66). This is what is referred to as the feature of normalisation, standardisation, conventionalisation or conservatism (Zanettin, 2012:19), which is defined as "the moulding of translations so they reflect what is routine in the target language" (Kenny, 1998:515).

As with the other features of translated language, normalisation can be studied from the perspective of S- or T-universals. As an S-universal feature, normalisation occurs when conventional target solutions are used to solve problems that are posed by unusual or creative source-text features, which in effect means that translations will contain fewer instances of creative or unusual target-language features than is to be expected in comparison with their respective source-language texts (Kenny, 2001:66). As such, normalisation manifests itself in a parallel corpus as an S-universal when translators use target-language solutions that are conventional for source-text problems that are creative in nature (Kenny, 2001:65). As a T-universal, normalisation occurs when translated texts contain more instances of a conservative feature that is considered typical of the target language compared to the frequency of the same feature in non-translated texts in the same language (Kenny, 2001:65).

Normalisation is also thought to contain an element of neutralisation, which refers to the opposition between common and less common collocations in the notion of collocational range (Øverås, 1998:12). Collocation refers to the patterning that arises when two or more words have a tendency to

occur near one another (Olohan, 2004:61). While collocations are usually understood as words that commonly occur near each other, reflecting habitual language use, collocations may also be unusual, idiosyncratic, creative or once-off usages (Kenny, 1998:521). Neutralisation occurs when the source text's creative collocational patterns are replaced with more habitual or common collocational patterns in the target language. According to Øverås (1998:569) who bases her discussion on the norm-oriented viewpoints of Toury, this means that translators are prone to using and producing structures or equivalents which are ready-made or pre-packaged and cliché in their translations. Baker (2007:14) argues that this is often the case with idioms in translation.

Neutralisation, however, is not the only aspect of normalisation which has a direct influence on the collocational patterns of the target text, as these can also be influenced by a process known as sanitisation. This refers to "the suspected adaptation of a source text reality to make it more palatable for target audiences" (Kenny, 1998:515). Sanitisation may become evident in investigating semantic prosody, which refers to the "consistent aura of meaning with which a form is imbued by itscollocates" (Louw, 1993:157), meaning that the sense of certain words is tainted by the words that typically surround them (Munday, 2011:170). Sanitisation may be evident when negative semantic prosodies in the source text are replaced by more neutral prosodies in the target text (Kenny, 1998:521) which would result in the target text having a more toned-down quality when compared to the source text (Kenny, 1998:520).

As is the case with the other features of translation, the concept of normalisation is not without contention. One of problems is that it may be difficult to draw the line between simplification and normalisation. Baker (1996), for example, originally suggested that simplification might be investigated by studying the occurrence of punctuation adjusted to target-language norms in translation, instead of the unusual punctuation of source texts – a method used by May (1997) to examine normalisation. Moreover, as stated above, simplification may be seen as related to the Unique Items Hypothesis, but normalisation appears to contradict this feature of translation (Castagnoli, 2008:23) since it is based on the assumption that translators use conventionalised translation equivalents, whereas the Unique Items Hypothesis postulates that translators closely keep to the make-up of source texts and, as a result, do not fully exploit the resources of the target language. This idea may again be linked to Toury's laws of standardisation and interference (see Section 2.3.2). Toury's laws also express the two directions in which translations are always pulled,

namely the effect of the source text, in the sense of interference, and conformity to target-language conventions, in terms of standardisation.

The concept of normalisation has also been linked to Toury's law of growing standardisation by Laviosa (2002:56) and Castagnoli (2008:20). In fact, Pym (2008) argues that all the features of translated language can "be more or less fitted" into the law of growing standardisation. His argument is centred on the idea that normalisation is linked to the concept of the selection of "habitual options" (Toury, 2012:304), simplification is the same as "reduced structuration" (Toury, 2012:203), and explicitation is contained in the metaphor of "flattening" (Toury, 2012:203). These correspondences lead Pym (2008) to claim that there are "some grounds for suspecting that all these universals are different aspects of one underlying universal". Yet, despite the evident validity of Pym's claim, his argument can be countered by the fact that some of the features of translated language can be traced back to source-text interference and are therefore related to Toury's law of interference – as pointed out above. In other words, even though some of the features are related to the law of standardisation, it should not be forgotten that they are also related to the law of interference, which complicates Pym's (2008) position.

(b) Results of previous studies

In general, studies of normalisation are based on the idea that the existence of this feature of translation will be evident in fewer instances of unusual or unattested usage in translated texts compared to their source texts, resulting in limited foreignness in translations (Yuan & Gao, 2008:3); or compared to non-translated texts in the same language. One of the most influential studies of normalisation as a possible S-universal has been done by Kenny (2001) who suggests that unusual, creative source-text words and collocations tend to be normalised in translation. Her investigation shows that in a German to English parallel literary corpus 44% of creative source-text lexical items were normalised in translation, as were 16% of creative collocations (Kenny, 2001). However, she also points out that there is evidence for originality and creativity in solutions to translation problems created by unusual source-text features. These solutions are therefore not necessarily always limited to normalisation, which can be interpreted to mean that there is rather ambiguous support for her hypothesis. Øverås (1998) finds evidence of similar shifts occurring in her corpus of Norwegian source texts and their target texts, and of English source texts and their translations, which results in

unusual collocations being turned into common collocations, and in the neutralisation of metaphorical expressions.

May (1997) has investigated normalisation by focusing on punctuation and how it is managed in translation. She concentrates on modernist experimental prose and finds evidence for normalising tendencies in Russian and French translations of William Faulkner and Virginia Woolf, particularly in terms of the syntax and the unusual punctuation of the texts. According to Olohan (2004:98) this can be interpreted as part of the tendency of translators to tone down or eliminate unusual features of the source text; however, intervention from editors cannot be ruled out.

Yuan and Gao (2008) investigate normalisation as a T-universal, and base their study on the hypothesis that normalisation will result in less "foreignness", fewer cases of unattested usage and a lower frequency of function words in translated texts compared to untranslated texts. The degree to which function words are used in relation to content words, in this instance, can be taken as indicative of the range of vocabulary used. They test their ideas by investigating lexical density, lemmas, and unattested or abnormal usage. Their study finds that there is a higher lexical density in their subcorpus of translated fiction than in the comparable subcorpus, which they take to indicate that translated texts are "even conscientiously more natural than non-translated texts in order to achieve higher popularity and acceptance among readers" (Yuan & Gao, 2008:10) – even though this runs contrary to Laviosa's (1998b:562) finding that lexical density is lower in translated narrative prose. In addition, they find evidence to support the notion that translators tend consciously to adopt idiomatic expressions so as to achieve the necessary equivalent effect of the source text (Yuan & Gao, 2008:22).

Williams (2005) studied normalisation by focusing on coinages, or "words with unattested orthography or atypical morphology" (Kenny, 2001:206). She searched for all forms of unattested words to be found in her translated and non-translated bidirectional corpus and found that both the English and the French translated subcorpora contained a considerably lower number of coinages than the corresponding non-translated subcorpora, even though they had a similar number of running words (Williams, 2005:104). She argues that the lower number of coinages in the translated subcorpora indicates a greater degree of conservatism and that translations may be normalised regardless of the languages involved (Williams, 2005:108).

Likewise, Kruger and Van Rooy (2012) studied the frequency of coinages along with loanwords. However, in their study no significant corpus effect was found for this specific operationalisation. In addition, their study focused on the frequency of lexical bundles, specifically trigrams, such as a number of, as well as, in terms of and in order to, as a reflection of conventionalised, "prefabricated" language use. They find a significant effect for register only, suggesting that the difference is between the registers of the two subcorpora rather than between the translated subcorpus and the non-translated subcorpus.

2.3.3.4 Levelling-out

(a) Overview

Baker (1996:184) defines "levelling-out" as the tendency of translated texts to "gravitate towards the centre of a continuum", and as such this feature means that translators typically steer a middle course between extremes. This implies that less variance will be encountered in terms of textual features, such as type-token ratio, lexical density and average sentence length (Yajun & Zaixin, 2008:29) in a corpus of translated texts than in a corpus of non-translated texts (Olohan, 2004:100). However, it should be pointed out that even though this particular feature of translated language has been noted in literature, it has been the subject of little empirical testing by translation theorists, and only a couple have commented on it (Williams, 2005:45). As a consequence, there is little evidence supporting the overall hypothesis (Olohan, 2004:100).

(b) Results of previous studies

Williams (2005) used readability indices as a means to measure levelling-out as a potential feature of translated language. A readability index refers to an instrument that is used in technical writing and in educational psychology to assess the degree of skill which is required to read a given set of texts. The purpose of using this instrument in this study, however, was not to determine the skill level required to read the texts in the subcorpora, but rather to discover whether the readability scores of the translated subcorpora were more homogeneous and, consequently, "levelled-out" (Williams, 2005:151). However, in this study, the results failed to consistently support the levelling-out hypothesis as it was found that contrary to expectations, the translated subcorpora required more advanced reading skills as a result of more complex sentence structures and vocabularies, and a greater frequency of optional syntactic elements (Williams, 2005:209). This finding is not only applicable to levelling-out, but also sheds more light on the difficult relationship between explication

and simplification. It shows that contrary to Pápai's (2004:160) claim that explicitation leads to lexical repetition which reduces the vocabulary range, explicitation by means of the inclusion of optional syntactic elements might actually increase the difficulty of translated texts as opposed to simplifying them.

Yuan and Gao (2008:22) hypothesised that the texts in a translation subcorpus would have a more harmonious set of scores (indicating greater "closeness", which would be expressed as lower standard deviation) in terms of sentence length, type-token ratio, lexical density and readability scores. They found only partial support for their hypothesis as their translated subcorpus showed greater homogeneity only in the case of type-token ratio, sentence length, and readability scores, and demonstrated greater dispersion of scores for lexical density (Yuan & Gao, 2008:18). For readability scores their study showed that the readability of the translated subcorpus varied little in comparison with the comparable subcorpus. This, in other words, indicates the relative homogeneity of the translated subcorpus. The implications of their study for levelling-out as a feature of translated language are that they thus found evidence for levelling-out in terms of sentence length, type-token ratio and readability scores but not for lexical density.

Kruger and Van Rooy (2012) operationalised levelling-out as a smaller degree of register variability in a translated subcorpus as compared to a non-translated subcorpus, occurring as a consequence of the effects of translation. They did not find overall support for their hypothesis as register differences were noted for all the linguistic features investigated irrespective of translated or non-translated status (Kruger & Van Rooy, 2012:60). Nevertheless, some subtle effects were noted: popular writing in their translational subcorpus did not appear to have same degree of informality as in the reference subcorpus, and translated academic writing seemed to be markedly pedantic in its usage of appositive linking adverbials (Kruger & Van Rooy, 2012:60).

2.4 Using the features of translated language as an index of translation expertise

It is evident that there is considerable support for the hypothesis that translated language differs from non-translated language, and that these differences manifest themselves in the features of translated language: explicitation, normalisation, simplification and levelling-out. Laviosa (2008a:307) believes that these features may occur as a consequence of three factors, namely the communicative role of translation, translators' awareness of their socio-cultural positions and roles, and the constrained

cognitive processing involved in the translation process. It is very likely that these three factors play varying roles for experienced and inexperienced translators. As discussed in Section 2.2, inexperienced translators differ from experienced translators in various respects. For example, they lack the translation competencies of experienced translators, experience different cognitive constraints, are less aware of the cognitive environments involved in translation, and fail to bridge the gap between procedurally, conceptually and contextually encoded information. As a consequence, the procedures, or translation strategies, used by inexperienced translators to solve problems are likely to be different from those used by experienced translators. It may thus be proposed that these differences in translation strategies will evidence in differences in the occurrence of the features of translated language in the translations of experienced and inexperienced translators.

2.5 Conclusion

This chapter provided background to and a context for the two concepts that inform the theoretical orientation and methodology for the study. The first part of the chapter explored the notion of translation expertise as it is currently understood in translation studies, while the second part concentrated on the concept of the features of translated language. The purpose of the discussion of these two sets of concepts was to investigate the implications of a possible relationship between them and to gain an understanding as to how or why the features of translated language may be taken as indicative of the translation processes and strategies used by translators, and consequently reflect translation expertise. In the next chapter, the methodology used will be discussed in more detail to provide more information on data collection and processing, the operationalisations of the features of translated language, and the hypotheses which inform the corpus analysis.

CHAPTER 3: METHODOLOGY

3.1 Introduction

The purpose of this study is to compare the incidence of the recurrent features of translation in translations produced by inexperienced and experienced translators. It is proposed that differences will be evident when their translations are compared to one another across various dimensions, and that these differences are the consequence of the different levels of expertise of the two groups. Specifically, it is hypothesised that the extent to which the features of translated language, such as explicitation, simplification, normalisation and levelling-out, occur in translations by the two groups may be seen as a reflection of the different translation strategies used by experienced and inexperienced translators.

In Section 3.2 an overview of the research design is presented, while Section 3.3 provides more information regarding the corpus composition. Section 3.4 focuses on data collection, while in Section 3.5 the operationalisations used to reduce the conceptual predictions to concrete, testable hypotheses are discussed in more detail. Section 3.6 outlines these hypotheses regarding the frequency with which linguistic realisations associated with each feature of translated language is expected to occur in the three subcorpora used for the study. Section 3.7 is devoted to an explanation of the statistical methods used to test the hypotheses, while Section 3.8 offers a conclusion to the chapter.

3.2 Research design

In terms of research design, a fundamental distinction between qualitative and quantitative research can be drawn. On the one hand, qualitative research aims to describe a quality of something in some enlightening manner and may focus on conclusions about what might happen, what is possible, or what could happen sometimes (Williams & Chesterman, 2001:64). As such, qualitative approaches are not suited to drawing conclusions concerning what is general, probable or universal (Williams & Chesterman, 2001:64). The purpose of quantitative research, on the other hand, is to formulate generalisations regarding a given phenomenon or feature, specifically in terms of how widespread or typical it is. Quantitative approaches therefore focus on frequencies, distributions, tendencies and regularities, and depend on the statistical comparison of figures (Williams & Chesterman, 2001:64-65). According to Williams and Chesterman (2001:65) many research projects contain elements of

both types of research: the qualitative approach is usually used in the initial stages where the categories and concepts needed for the study are set up and defined, whereas the quantitative approach is used during the analysis stage.

Corpus-based research utilises a large, principled collection of texts, known as a corpus, as the basis for the analysis. Corpus-based methods are empirical in that linguistic patterns are analysed as they occur in natural texts. Computers are used extensively both for interactive and automatic data-collection techniques (Biber *et al*., 1998:4). Corpus-based research may engage qualitative as well as quantitative methods. In the context of corpus-based translation studies, Munday (2009:181) points out that translation research based on a corpus methodology may be both qualitative, in that the concordance lines which contain individual instances are analysed, and quantitative, in that numerical data for a variety of specific and general linguistic and textual features are subjected to statistical processing and interpretation. The basic approach of this study is quantitative. Hypotheses are formulated, and corpus data are analysed quantitatively in order to test these hypotheses. Qualitative analysis configures the categories and features for analysis.

The corpus for this study consists of texts written in a single language, English, and consequently constitutes a monolingual corpus. A comparable corpus design was used which consists of separate sets of text collections in the same language. These separate text collections consist of two translational sets, composed of target texts produced by inexperienced translators and experienced translators, respectively, and a non-translational set used as a reference corpus. This design allows one to determine whether there is a difference in the patterning of translations produced by experienced and inexperienced translators in terms of the hypothesised recurrent features of translation. The reference corpus consisting of non-translated English was included to allow a comparison which would determine if the translation features identified in the translations of experienced and inexperienced translators are indeed particular to translated texts, or if they also occur to the same degree in original, non-translated texts.

A couple of considerations need to be kept in mind concerning corpus-based research on translation that utilise comparable corpora. The notions of comparability and representativeness are essential considerations in corpus-based studies (Olohan, 2004:42). Comparability means that the text collections should be as similar in as many respects as possible in order to ensure that their

characteristic linguistic differences can be attributed to their non-translated or translated status (Laviosa, 2002:39), or, in the case of this study, to the relative differences in experience of the two groups of translators involved. For this reason, it is important to aim for comparability between corpora along different dimensions, which includes the text or register categories, the overall size of the corpora, the time span of when the texts were published or produced, as well as the distribution of male and female authors, along with team or single authorship (Olohan, 2004:42).

In addition to comparability, Halverson (1998:3) argues that a careful description of what the corpus is intended to represent is also necessary so as to ensure that the data are representative of a particular genre, language, text purpose, or text function. Furthermore, temporal and regional criteria may need to be taken into account and can include factors such as nationality, age, gender or ethnicity of the writers or the translators (Olohan, 2004:46). The establishment of the particular criteria will be determined by the hypotheses to be tested, the research questions which are to be addressed, and the aim of the research (Olohan, 2004:46). On account of the importance of the notions of comparability and representativeness, the methodological foundations for the compilation of the three subcorpora used in the study will first be discussed.

3.3 Corpus composition

3.3.1 Compilation

In order to answer the research questions of this study, it was necessary to compile a comparable corpus consisting of three custom-built subcorpora, representing translations produced by inexperienced translators, translations produced by experienced translators and non-translated texts. The three subcorpora have been named the inexperienced translators' subcorpus (IT), the experienced translators' subcorpus (ET) and the non-translated subcorpus (NT). It is important to point out that because of its planned use as a comparable corpus, the ET and NT subcorpora were constructed to match the IT subcorpus. The main reason for this is the difficulty encountered in collecting translations which were not produced by professional, experienced translators.

The translations selected for the IT subcorpus were produced by either student translators or laypersons – translators who have little or no experience in professional translation. Some of the student translators, who translated from Afrikaans to English, were enrolled for a translation programme at the North-West University, University of Johannesburg, Stellenbosch University, and

University of the Free State. These students generally had Afrikaans or English as a first language. In a couple of instances the source texts of translations from some of the South African universities were French, German or Mandarin Chinese. Other student translations include translations from Polish into English, consisting mostly of legal texts, by students whose first language is Polish studying at the Adam Mickiewicz University in Poznan. The majority of the student translations were produced to be marked as class assignments. The student translators were all in their first to fourth year of studies.

To represent translations done by laypersons, texts were selected from the Internet. In all these cases the writer specified that the text is a translation. In a number of instances they clearly indicated that they were not professional translators or that they had little translation experience, but produced the translations for fun or with the intention of getting feedback from native English speakers. In some cases they indicated that they did the translation because English readers were excluded from the readership of the text as it had not been officially translated into English despite the popularity or high standing of the source text in its native country. This was particularly the case for the Japanese creative writing source texts. Texts where the translator specified that they do not speak the source language and only managed to translate the text by means of translation tools, such as Google Translate, were automatically excluded. French, Spanish, German and Swedish were the main source languages of the translations produced by laypersons.

The texts contained in the ET subcorpus were collected from printed published works (which were scanned and converted into a machine-readable format by means of an optical character recognition programme), from the Internet, or from an existing translation corpus. The printed published works mainly consisted of book samples of which a number form part of the academic register, with Italian, French and German as source languages. The Internet texts comprise legal documents and magazine and newspaper articles, from a variety of source languages including Spanish, French, Danish and Dutch. The remainder of the texts were selected from a corpus compiled for another study (see Kruger & Van Rooy, 2012), and are all texts translated from Afrikaans into English. In general, it was assumed that these texts are instances of authentic, naturally occurring translations which were realised as a consequence of actual translation needs and were produced by professional, experienced translators as they were published or used in corporate environments.

The texts contained in the NT subcorpus were collected in a similar manner to the ones contained in the ET subcorpus. Some of the texts were collected from books which were scanned before an optical character recognition programme was used to convert the text into an electronic format. Others were copied from South African, British, and American Internet sites of popular magazines and newspapers. These sites were selected because they reflect the standard written form as sanctioned by publishing industries. In addition, a number of texts were selected from the British National Corpus (BNC), which contains extracts of British texts. Texts were also taken from the International Corpus of English for South Africa (ICE-SA) to match the representation of South African English in the IT subcorpus and the ET subcorpus. Since the two translation subcorpora include translations produced in a variety of contexts, it was deemed necessary to include American, British, and South African English in the NT subcorpus.

All the texts in the corpus were produced in a 30-year period from 1982 to 2012. This restricted time frame means that the corpus constitutes a synchronic corpus. As comparability is an important consideration in corpus compilation, effort was made to ensure that the corpus was as balanced as possible in terms of register, word count and, in the case of the two translation subcorpora, source languages. These three dimensions are discussed in the next section.

3.3.2 Registers, word count and source languages

The term "register" refers to the language varieties that are associated with different purposes and circumstances as defined by their various situational characteristics (Biber *et al*., 1998:135). For this study, the written text categories used by the ICE were used to classify texts by register for all three subcorpora, namely: academic, creative, instructional, popular, reportage, and persuasive writing. Academic writing can be considered as texts which "were written by academics for academics" and sources include journal articles or academic books (Nelson *et al*., 2002:7). A variety of topics may be included which range from history, literary criticism and economics to astronomy, robotics and surgery (Nelson *et al*., 2002:7).

Creative writing is constituted by narrative or descriptive prose and dialogue passages which encompass a number of fiction genres, such as thrillers, science fiction and detective novels (Nelson *et al*., 2002:8). Instructional writing, according to criteria set out by the International Corpus of English for Great Britain (ICE-GB), is divided between regulatory or administrative writing and

publications which deal with hobbies and skills (Nelson *et al*, 2002:8). The regulatory/administrative category is corporate in nature and is written on behalf of government or other administrative bodies with the aim of transmitting information to the general public (Nelson *et al*., 2002:8). Texts which form part of the hobbies/skill component offer instruction which is directed towards a more specialised readership and may include cookery books and gardening manuals (Nelson *et al*., 2002:8).

Popular writing, in turn, responds to the needs and requirements for accessible reading materials of a broad literature public and is written with the general public in mind (Biber, 1995:300). It typically includes magazine articles written for a general audience. Reportage includes general domestic and foreign news, leading articles, business reports, and sport reports. In general, the reportage component is distinguishable from press editorials as the intention of news reports is to inform rather than to persuade (Nelson *et al*., 2002:8). The persuasive writing category, in other words, is less directly related to current news and contains texts that provide the writer with the opportunity to be discursive in a way which news journalism does not (Nelson *et al*., 2002:8).

As very few translated texts could be obtained for the persuasive writing register, it was decided to remove this register from the list and to categorise the texts it contained into other suitable register categories. The final word count for the five registers in the three subcorpora is displayed in Table 3.1.

	NT	**IT**	**ET**
Academic writing	33939	32310	33504
Creative writing	33018	30462	29840
Instructional writing	25928	34039	26133
Popular writing	36768	40381	37482
Reportage	28910	26448	28818
Total	**158563**	**163640**	**155777**

Table 3.1: Word counts by register in the three subcorpora

Even though the corpus designed for this study is quite small in terms of the word count, Becher (2011:30) is of the opinion that in studies such as these where the texts were produced by different authors or translators, reliability should not be a problem if the lexical material is spread over a large number of texts. In the case of this study there are 84 texts in the NT subcorpus, 85 translations in the

IT subcorpus and 87 translations in the ET subcorpus. The texts for the entire corpus were either extracts or full texts. As the goal was to work with units of texts which were no shorter than 1000 words so as to avoid the potential analysis problems that are associated with shorter texts, shorter texts of the same text type were combined to create a text longer than 1000 words. Each of the texts contained in the corpus includes header information which provides more information on the text.

The NT subcorpus consists mainly of examples of British and South African English with some American English interspersed. British and American English were included since they are regarded as global standard varieties especially in the international publishing industry, and South African English was included to match the representation of South African English in the two translation subcorpora. Other Englishes were not included, in order to limit other sources of variability in the NT subcorpus. The overall word count and percentage of the three varieties in the NT subcorpus in terms of the five different registers are displayed in Table 3.2.

Register	**American English**		**British English**		**South African English**	
	Word count	Percentage of corpus	Word count	Percentage of corpus	Word count	Percentage of corpus
Academic writing	5211	15.35	15741	46.38	12987	38.27
Creative writing	6562	19.87	20072	60.79	6384	19.33
Instructional writing	6384	24.62	11127	42.91	8417	32.46
Popular writing	11159	30.35	13479	36.66	12130	32.99
Reportage	7080	24.49	11773	40.72	10057	34.79
Total	**36396**	**22.94**	**72192**	**45.49**	**49975**	**31.57**

Table 3.2: Representation of varieties of English in the NT subcorpus for the five registers

As can be seen from Table 3.2, American English constitutes 22.94%, British English 45.49% and South African English 31.57% of the NT subcorpus.

In the academic writing category of the NT subcorpus, the American English texts deal with medical, scientific and historical topics. The British texts focus on history, psychology and scientific topics

whereas the South African texts are concerned with computers and education. Both the British and South African components also feature texts on geography and linguistics. In the creative writing category, the American texts are samples of best-selling novels written by Dan Brown, Jonathan Kellerman, Michael Crichton and Nicolas Evans. The British and South African fiction texts included, however, are not quite as main-stream. The British texts are mostly extracts from detective and suspense novels whereas the South African texts are satirical in nature.

The texts in the instructional category in the NT subcorpus are all similar to one another in that they are mostly regulatory or corporate in nature and have been written by administrative bodies with the intent to transfer information to either the general public, as with legal legislation acts, or to convey company information to employees of a particular business. In the popular writing register, the American English texts deal with diving, computers and various humoristic topics whereas articles on travel, cuisine and celebrities constitute the British component. The South African texts focus on parenting, cars, travel and local celebrities.

In the reportage category of the NT subcorpus, texts mainly comprise of online newspaper sources such as The New York Times, The Washington Post and The Dallas Morning News for the American English component; Scotland Herald and Daily Telegraph for the British constituent; and News24 along with some newspaper texts taken from ICE-SA for the South African component. For all three of the English varieties in the register of reportage, the articles focused primarily on business and commerce, and local and international news.

Although the overall word counts in Table 3.1 indicate that the corpus is relatively balanced, and a similar number of texts are included in the three subcorpora, it was much more difficult to balance the corpus in terms of the representation of different source languages in the different registers in the two translation subcorpora. It will become evident in the discussion that follows that some source languages are overrepresented while others are underrepresented in some of the registers.

Translations from a number of source languages including the Germanic, Romance, Altaic, Slavonic, Sino-Tibetan, Finno-Ugric, Indic and Semitic families were included. The word count and the percentage of the source languages in the two translation subcorpora are presented in Table 3.3.

	IT		ET	
Language	Word count	Percentage of corpus	Word count	Percentage of corpus
Germanic				
Afrikaans	56067	34.26	56234	36.11
Danish	1003	0.61	5996	3.85
Dutch	4188	2.56	11007	7.07
German	6631	4.05	6391	4.10
Norwegian	5838	3.57	5298	3.40
Swedish	11077	6.77	6224	4.00
Romance				
French	12854	7.86	17728	11.38
Italian	1751	1.07	1610	1.03
Spanish	15206	9.29	9258	5.94
Altaic				
Japanese	12357	7.55	6138	3.94
Korean	1176	0.72	2336	1.50
Slavonic				
Polish	14377	8.79	3499	2.25
Russian	5413	3.31	3898	2.52
Bulgarian	1013	0.62	0	0.00
Sino-Tibetan				
Chinese	13384	8.18	13249	8.51
Thai	0	0.00	1393	0.89
Finno-Ugric				
Estonian	0	0.00	1501	0.96
Indic				
Nepali	0	0.00	2854	1.83
Semitic				
Arabic	1305	0.80	1128	0.72

Table 3.3: Representation of source languages for the two translation subcorpora

As is evident in Table 3.3, the Germanic source languages are strongly represented in both the translation subcorpora, especially in terms of Afrikaans, which contributes to 34.26% and 36.11% of the IT and ET subcorpora respectively. This overrepresentation is the consequence of the fact that the majority of the student translations collected for the IT subcorpus have Afrikaans as source language. This bias is consequently reflected in the ET subcorpus, which was designed to be as comparable as possible to the IT subcorpus. In the NT subcorpus this bias is matched by the representation of South African English, which constitutes 31.57% of this subcorpus (see Table 3.2). In the case of some source languages, such as Arabic, Nepali and Bulgarian, a language was represented in one of the

subcorpora but not the other because of the difficulty involved in tracking down and collecting translations done in these languages.

Tables 3.4 to 3.8 show the word counts and the overall percentage of source languages represented for each register across the subcorpora.

Language	IT		ET	
	Word count	Percentage of corpus	Word count	Percentage of corpus
Germanic				
Afrikaans	20036	62.01	19076	56.94
Danish	1003	3.10	1130	3.37
Dutch	1258	3.89	1117	3.33
German	1444	4.47	1443	4.31
Swedish	1259	3.90	1181	3.52
Romance				
French	1982	6.13	2136	6.38
Italian	1751	5.42	1610	4.81
Spanish	1099	3.40	1073	3.20
Altaic				
Japanese	0	0.00	1176	3.51
Slavonic				
Russian	1195	3.70	2359	6.90
Sino-Tibetan				
Chinese	1283	3.97	1203	3.59

Table 3.4: Representation of source languages in the two translation subcorpora for the academic register

Looking at the figures for the academic register (see Table 3.4), it is evident that the Germanic languages are strongly represented, mostly attributable to translations done from Afrikaans which constitutes 62.01% of the IT subcorpus and 56.94% of the ET subcorpus. The translations from other languages constitute 3 to 6% each in the first subcorpus and 3 to 7% each in the second. The academic texts of the IT corpus are study guide translations, as well as translations submitted to Wikipedia. The study guide translations have Afrikaans as source language, and lecturers as well as students as target audience. The Wikipedia translations are from a variety of languages, and were submitted to Wikipedia together with a request from the translators for feedback from English speakers. The study guide and Wikipedia translations are used to represent the academic register in the IT subcorpus, even though these types of texts are perhaps not typical academic texts. It appears

that inexperienced and particularly lay translators infrequently self-select academic texts for translation – an assumption made on the basis of the difficulty encountered in collecting such translations.

The study guide translations are mainly from the fields of Communication Studies and Mathematics, whereas the Wikipedia articles are mainly of a historical, encyclopaedic nature. The ET subcorpus is composed mostly of textbook translations, including philosophical, medical and historical subjects. In general, the representation of source languages in the two subcorpora is comparable for the academic register.

Language	**IT**		**ET**	
	Word count	Percentage of subcorpus	Word count	Percentage of subcorpus
Germanic				
Afrikaans	1983	6.51	2030	6.80
Dutch	0	0.00	2644	8.86
Norwegian	2871	9.42	2819	9.45
Swedish	8170	26.82	3179	10.65
Romance				
French	1471	4.83	11492	38.51
Altaic				
Japanese	10593	34.77	1673	5.61
Slavonic				
Polish	1602	5.26	1020	3.42
Russian	2025	6.65	1539	5.16
Sino-Tibetan				
Chinese	1747	5.74	3444	11.55

Table 3.5: Representation of source languages in the two translation subcorpora for the creative writing register

In the creative writing register (see Table 3.5), translations with Japanese and Swedish source languages are overrepresented in the IT subcorpus, constituting 34.77% and 26.82% of the subcorpus, respectively. These percentages are not matched in the ET subcorpus where translations from these languages represent 5.61% and 10.65% of the subcorpus. French, on the other hand, is strongly represented in the ET subcorpus (38.51% of this subcorpus), but target texts from this language represent only 4.83% of the translations produced by inexperienced translators. The imbalance is caused by the difficulty encountered in the collection of texts that are produced by inexperienced

translators in this register, with the exception of translations with Japanese source texts. These are relatively widespread on the Internet because of the influence of Japanese popular culture, with numerous lay translators opting to translate texts with anime and manga roots.

It proved difficult to collect similar translations produced by experienced translators and, in order to balance the two subcorpora in terms of the number of words in this register, another source language was used to even the word count. Translations with French as a source language done by experienced translators were easy to collect, which accounts for its overrepresentation in the ET subcorpus. In addition to the texts with anime sources, a number of the translations in the IT subcorpus are from the fantasy genre. In the ET subcorpus a wider range of fictional genres are present, which include, but are not limited to; fantasy, thrillers, and detective novels.

Language	**IT**		**ET**	
	Word count	Percentage of subcorpus	Word count	Percentage of subcorpus
Germanic				
Afrikaans	10298	30.25	9594	36.71
Danish	0	0.00	1604	6.14
Norwegian	0	0.00	1242	4.75
Romance				
French	0	0.00	1300	4.97
Spanish	7760	22.80	4166	15.94
Slavonic				
Polish	12775	37.53	2479	9.49
Russian	2193	6.44	0	0.00
Bulgarian	1013	2.98	0	0.00
Sino-Tibetan				
Thai	0	0.00	1393	5.33
Finno-Ugric				
Estonian	0	0.00	1501	5.74
Indic				
Nepali	0	0.00	2854	10.92

Table 3.6: Representation of source languages in the two translation subcorpora for the instructional register

As is evident in Table 3.6, it was difficult to balance the representation of source languages in the instructional writing register, since it proved challenging to obtain translations from the same source languages for both subcorpora. The Polish and Afrikaans translations in the IT subcorpus were

provided by some of the universities mentioned in Section 3.3.1 and, as such, are produced by students. The translations from Afrikaans are administrative or informative in nature whereas the Polish translations all deal with legal proceedings, specifically notary deeds. The other languages represented in this register in the IT subcorpus had to be collected from the Internet, which was made difficult owing to the fact that it appears that instructional texts are not a preferred text type for lay translators, who tend to be more interested in popular and creative texts. However, some legal translations done from Spanish, Bulgarian and Russian which deal with administrative travel arrangements were included in the subcorpus.

For the ET subcorpus, some examples of translations of an administrative nature were collected from the Internet, but these often do not have the same source languages as the texts in the IT subcorpus. Some of these translations provide administrative information on local events or registration procedures for an organisation, especially in the case of the Afrikaans source texts, whereas the other texts are mostly translations of legal acts, primarily from Spanish, Nepali, Danish and Norwegian.

Language	**IT**		**ET**	
	Word count	Percentage of subcorpus	Word count	Percentage of subcorpus
Germanic				
Afrikaans	21310	52.77	20052	53.50
Danish	0	0.00	3262	8.70
Dutch	0	0.00	2971	7.93
German	3766	9.33	1004	2.68
Norwegian	1175	2.91	0	0.00
Romance				
French	6491	16.07	1252	3.34
Spanish	2188	5.42	1712	4.57
Altaic				
Korean	0	0.00	1197	3.19
Sino-Tibetan				
Chinese	5451	13.50	6032	16.09

Table 3.7: Representation of source languages in the two translation subcorpora for the popular writing register

In terms of popular writing (see Table 3.7), it is evident that the Germanic languages are once again overrepresented, specifically with Afrikaans, which represents 52.77% and 53.50% of the IT and ET subcorpora respectively. It is also evident that some of the languages are represented in either the IT

or ET subcorpus, but not in both. This is once again caused by the fact that it was difficult to source translations in this register from the Internet. However, the popular texts in the two translational subcorpora are quite similar, and include interviews, popular biographies or event reviews. In the IT subcorpus, the translations from Afrikaans are all class assignments and deal with a range of topics from car advertisements to movie or event reviews to short biographies about celebrities. The Internet texts are interviews with celebrities (especially in the case of the texts translated from French and Spanish), or texts that focus on technological developments (mostly in the case of translations with German or Norwegian as source languages). The Chinese source language text is a written recollection of experiences of a rural Chinese town.

The Afrikaans source language texts contained in the ET subcorpus are similar to the ones in the IT subcorpus in that they mostly focus either on reviews of popular South African events or are popular biographies of important figures in the country. Others provide information on topics such as viticulture and the cultivation of mushrooms. There are also online magazine articles consisting of interviews with or biographies about influential individuals (for translations with Korean, Danish, Chinese and Dutch as source language), and cuisine-related articles, with primarily French and German source texts.

Language	IT		ET	
	Word count	Percentage of subcorpus	Word count	Percentage of subcorpus
Germanic				
Afrikaans	2440	9.23	5482	19.02
Dutch	2930	11.08	4311	14.96
German	1421	5.37	3944	13.69
Norwegian	1792	6.78	1237	4.29
Swedish	1648	6.23	1864	6.47
Romance				
French	2910	11.01	1547	5.37
Spanish	4159	15.73	2307	8.01
Altaic				
Japanese	1764	6.67	3289	11.41
Korean	1176	4.45	1139	3.95
Sino-Tibetan				
Chinese	4903	18.55	2570	8.92
Semitic				
Arabic	1305	4.93	1128	3.91

Table 3.8: Representation of source languages in the two translation subcorpora for the reportage register

Looking at the figures for the reportage register (see Table 3.8) it is evident that even though both subcorpora contain extracts of the same languages, the percentages of the subcorpus these languages constitute are not necessarily similar. Translations from Chinese are predominant in the IT subcorpus at 18.55%, whereas translations from Afrikaans dominate in the ET subcorpus, at 19.02%. It should be noted that a number of texts contained in the IT subcorpus might actually have been better suited to the persuasive writing register since they are specifically translated to create an awareness of difficulties faced by particular countries – whether these be religious, financial or political. However, as there were not enough texts to sufficiently represent the register of persuasive writing, these texts were used to augment the reportage register, as they do refer to newsworthy events. This is the case with the translations from Norwegian, Dutch and Spanish. The other texts are mainly newspaper reports. For the ET subcorpus, the majority of texts are taken from online newspapers, and the texts focus on current national and international events, and financial topics. The only texts in the ET subcorpus for this register which were not collected from the Internet are those with Afrikaans as source language, but the texts are comparable to the others contained in the ET subcorpus in that they have the same scope and focus on similar types of events.

3.4 Data collection

Once the corpus had been compiled as described above, some preliminary processing of the texts was necessary to facilitate data collection.

3.4.1 Part-of-speech tagging

According to Olohan (2004:51) it may be necessary to perform some processing on the texts once they have been converted to a machine-readable format before they are analysed with corpus software. The extent of the text processing is determined by a number of factors which include the purpose of the corpus and the corpus software which will be used (Olohan, 2004:51). Annotated tags are an example of text processing which can be useful for some corpus-based studies. The most common type of annotated tag is one which indicates what part of speech a particular word is, which is known as grammatical tagging or part-of-speech (POS) tagging. According to Zanettin (2011:120) the advantages of adopting an annotation framework are, amongst others, that it can be automated to an extent, it is possible to ignore existing tags in a corpus or even to build on existing annotation, and it allows for a broader variety of investigations than just plain text.

For the purposes of this study, it was considered necessary to use a tagged corpus to collect data for some of the operationalisations (see Section 3.5). The POS-tagging software selected to tag the texts is known as CLAWS (the Constituent Likelihood Automatic Word-tagging System) developed at Lancaster by UCREL. This software boasts a consistent achievement of 96% to 97% tagging accuracy, depending on the type of text, which is considered to be reasonable for automatic tagging (Olohan, 2004:52). The most recent version of the tagger (CLAWS4) has also been used to POS-tag the British National Corpus (BNC)[7].

3.4.2 Data collection

WordSmith Tools 5.0 corpus-analysis software (Scott, 2008) was used for the extraction and collection of the data. WordSmith Tools is widely used for corpus-based studies in a variety of different languages with various aims. The program has three components. The WordList tool

[7] The British National Corpus is a corpus that consists of 100 million word collections of spoken and written English from a range of sources that is designed to represent current British English. More information is available at http://www.natcorp.ox.ac.uk/. (Date of access: 30 April 2013.)

provides the researcher with a list of all the different words or word-clusters in selected texts which can be set out in frequency or alphabetical order (Scott, 2008). KeyWords gives the researcher the opportunity to see all the key words in the texts selected, whereas the concordancer, Concord, allows the researcher to see any phrase or word in context (Scott, 2008).

For each of the features of translated language discussed in Chapter 2, a selection of linguistic operationalisations was chosen, usually on the basis of existing research in the field. These operationalisations are outlined in detail in Section 3.5. For each operationalisation a hypothesis was formulated indicating how the dependent variable (the feature under investigation) would vary by the independent variable of corpus (IT, ET and NT) (see Section 3.6). WordSmith Tools was used to extract the data necessary for each operationalisation. The data were then analysed and explained in Section 3.7 in order to test the hypothesis.

The data collection process depends on the clear definition of concrete linguistic phenomena associated with the abstract categories of the features of translated language – in other words, the operationalisation of concepts such as explicitation, normalisation, simplification and levelling-out in linguistic terms. The following section provides more detail regarding the operationalisations used, and the exact data extraction terms and strategies used to collect data for each feature of translated language investigated in this study.

3.5 Operationalisations

3.5.1 Explicitation

3.5.1.1 *That*-omission (E1 operationalisation)

As discussed in Chapter 2, explicitation in English literary texts has been investigated by Olohan and Baker (2000), who operationalise this feature of translated language in terms of the patterns of omission and inclusion of the optional *that* occurring with the reporting verbs TELL and SAY. Since the reporting *that* is an optional or redundant syntactic element, its inclusion can be seen as a linguistic manifestation of explicitation in translation if it is included more frequently in translations than comparable non-translations even though there might be "no obvious justification for doing so" (Olohan, 2001:424). Other studies which have used this methodology, with various other reporting verbs, include Williams (2005), Olohan (2004) and Kruger and Van Rooy (2012). For this study, the list of reporting verbs provided by the Cobuild English Grammar (1990:270) was used as search

terms for the investigation (see Table 3.9). In order to do so, a concordance for all forms of these verbs in the three subcorpora was produced by means of WordSmith Tools. Irrelevant entries were manually discarded and concordance lines were sorted and tagged to identify instances of *that*-omission and inclusion. Relevant entries were typically such as the following taken from the ET subcorpus where the optional *that* is omitted: "[…] because they assumed it would offend their readers." An example of an irrelevant entry also taken from the ET subcorpus reads: "If you studied, you could assume the office." *That*-omission was analysed as a ratio of all possible instances where a choice between the two forms could have been made.

accept	consider	gather	observe	report
acknowledge	contend	guarantee	perceive	resolve
add	convince	guess	persuade	reveal
admit	decide	hear	pledge	say
agree	deny	hold	pray	see
allege	determine	hope	predict	sense
announce	discover	imagine	promise	state
answer	dispute	imply	prophesy	suggest
argue	doubt	inform	read	suppose
assert	dream	insist	realis(z)e	swear
assume	elicit	judge	reason	teach
assure	estimate	know	reassure	tell
believe	expect	learn	recall	think
boast	explain	maintain	reckon	threaten
claim	fear	mean	record	understand
comment	feel	mention	reflect	vow
complain	figure	note	remark	warn
concede	find	notice	remember	wish
conclude	foresee	notify	repeat	worry
confirm	forget	object	reply	write

Table 3.9: List of reporting verbs

3.5.1.2 Conjunctive markers (E2 operationalisation)

Mutesayire (2004) has investigated reformulation, marked by apposition markers, as a kind of explicitation. The study is based on the assumption that the use of apposition might flag instances of explicitation as a discourse process since apposition involves the restatement of previously given information in a more explicit form so as to minimise ambiguity or to help with the interpretation of the text. Apposition markers, in other words, are taken to be indicative of translators' tendency to make the links between propositions clearer in the text.

This study, however, was not just limited to apposition markers, but focused on conjunctive markers more broadly, including the three types of conjunction distinguished by Halliday and Matthiessen (2004:541), namely: elaboration, extension and enhancement, which are each divided into particular subtypes. These types of conjunctives will be discussed in more detail along with the tables that follow (Table 3.10 to 3.12). Conjunctions can be seen as indicators of explicitation since they mark a number of different ways that explication can occur as a discourse process. These include the repetition of information previously given to avoid ambiguity, the specification of a reference, the addition of explanatory phrases, the expansion of condensed passages, and the addition of cohesive devices to enhance the flow of the text. Against this background, it should be noted that this operationalisation and the E1-operationalisation (*that*-omission) are indicators of different kinds of explicitation. Whereas the E1-operationalisation is concerned with explicitation as a kind of syntactic redundancy, the E2-operationalisation deals with explicitation as a discourse process.

For this study, concordance lists were produced for markers typically indicating the three different types of conjunction in the three subcorpora. Some manual sorting was required to remove irrelevant entries, and to correctly categorise some conjunctions, such as however and then, which may have more than one function and may therefore belong in more than one category of conjunction. An example of a relevant entry from the NT subcorpus reads: "[…] all societies appear as social spaces, that is, as structures of differences […]" where the phrase that is is used to introduce an explanatory phrase. An example of an irrelevant entry reads: "That is not the case with the topographical plan […]".

Elaboration takes place when clauses elaborate on the meaning of other clauses by further describing or specifying them (Halliday & Matthiessen, 2004:396), and its subtypes consist of apposition and clarification. Apposition is a type of elaboration where an element is restated or represented either by exposition or by example (Halliday & Matthiessen, 2004:540). Clarification occurs when the elaborated element is summarised, restated or made more precise (Halliday & Matthiessen, 2004:541). For a list of the elaboration markers used for this study refer to Table 3.10.

Elaboration	
Apposition	**Clarification**
e.g.	actually
for example	anyway
for instance	as a matter of fact
i.e.	as I was saying
I mean	at least
I mean to say	briefly
in other words	by the way
namely	in any case
that is	in conclusion
that is to say	in particular
to be exact	in short
to be more exact	incidentally
to be more precise	leaving that aside
to be precise	more especially
to illustrate	or rather
to put it another way	to be more precise
	to get back to the point
	to resume
	to sum up

Table 3.10: List of elaboration markers

A conjunctive relationship of extension exists when clauses extend the meaning of other clauses by adding something new to them. It consists of addition and variation (Halliday & Matthiessen,

2004:405). Addition may be positive, negative or adversative in nature, while variation includes subtractive, replacive and alternative types (Halliday & Matthiessen, 2004:543). The list of conjunctive markers of extension used as search terms in the study is provided in Table 3.11. The word and, which forms part of the addition category, was not included, as it would have required an unmanageable amount of manual sorting because of its ambiguous nature as a clause coordinator and phrase coordinator. For example, an example of and as a clausal connector taken from the IT subcorpus reads: "I quickly straddled the bike and I sat down […]" whereas an example of and as a phrasal coordinator from the IT subcorpus reads: "Sasaki, Kunikida, and myself gazed at the students […]". In addition, further reason to exclude and is because its frequency tends to be more than all other conjunctions, even when used as a clausal coordinator only, which means that the findings are not the results of the elaboration category, but rather of the word itself [8].

Extension	
Addition	**Variation**
also	alternatively
moreover	apart from that
in addition	except for that
nor	instead
but	on the contrary
on the other hand	or else

Table 3.11: List of conjunctive markers of extension

Conjunction involving enhancement takes place when clauses enhance the meaning of other clauses by qualifying them in various ways. Four subcategories of enhancement conjunctions are distinguished: causal-conditional, manner, matter, and spatio-temporal (Halliday & Matthiessen, 2004:210). In the causal-conditional category, the relation of cause figures as a cohesive agent, whereas manner conjunctives establish cohesion either by comparison or by reference to means (Halliday & Matthiessen, 2004:546). Matter conjunctives create cohesion by reference to the preceding "matter", and spatio-temporal conjunctives mark the use of spatial relations as "text-creating cohesive devices" (Halliday & Matthiessen, 2004:544-547). An example of a causal-

[8]For explicit treatment of this issue, refer to Van Rooy and Esterhuizen (2011).

conditional conjunctive is found in the following: “Still, many foreign investors say they believe in the principle […]” while an instance of a spatio-temporal conjunctive occurs in the following: “Finally, a matrix of rank can be expressed […]”. Both examples are taken from the NT subcorpus. The enhancement conjunctions used as search terms in this study are presented in Table 3.12.

Enhancement			
Causal-conditional	**Manner**	**Matter**	**Spatio-temporal**
all the same	by such means	as to that	after a while
as a result	in a different way	elsewhere	afterwards
as to that	likewise	here	all that time
because of that	similarly	in other respects	at once
consequently	thereby	in that respect	at the same time
despite this	thus	there	at this moment
even so			at this point
for			before that
for that purpose			finally
for that reason			hitherto
hence			in the end
here			just then
if not			last of all
in consequence			lastly
in other respects			meanwhile
elsewhere			next
in that case			next time
in that respect			now
nevertheless			on another occasion
on account of this			previously
otherwise			secondly
still			soon
there			straightaway
therefore			thereupon
though			until then
under the circumstances			up to that point
with this in view			

Table 3.12: List of enhancement conjunctions

3.5.2 Simplification

3.5.2.1 Type-token ratio (S1 operationalisation)

Type-token ratio (TTR) refers to the ratio of unique word forms (or types) to the running words (or tokens) in a text, and is used as a measure of vocabulary range, or lexical diversity, in a text (Kenny, 2001:34). It also provides a "simple measure of the superficial lexical complexity of a text" in that the higher the TTR is, the more varied the vocabulary is (Munday, 1998:4). Conversely, a text with a lower TTR has a smaller vocabulary range, and may therefore be regarded as simpler, in this lexical dimension, than a text with a higher TTR. Based on this, TTR may therefore be used as an operationalisation to investigate simplification, since lexical diversity as measured by TTR corresponds with complexity at the level of vocabulary range.

Since TTR is very sensitive to text length, the ratio was standardised for it to be of use. In this study, TTR was standardised for 1000 words, which means that the TTR was calculated anew for every 1000 words and was then averaged for the entire text. These calculations were performed by WordList in WordSmith Tools 5.0.

3.5.2.2 Readability score (S2 operationalisation)

Readability indices, according to Williams (2004:72), are instruments that measure the degree of complexity of the syntax and morphology of written texts before the results are correlated with the reading skill level of a particular population. These instruments, in other words, assess the degree of skill needed to read a particular text. The readability indices are measured on a conceptual continuum which ranges from "easy to read" to "hard to read" (Williams, 2004:72). As such, readability scores may provide an operationalisation of simplification that is more syntactical and morphological in nature than the S1 operationalisation, which is more lexical. Examples of readability indices include Björnsson's LIX, the Gunning Fog Index, the Fry Readability Graph, the Flesch Reading Ease Test and the Flesch-Kincaid Grade Level Test.

Sentence length, which is an important parameter for reading indices, has been used by Laviosa (1998a, 1998b) to investigate simplification, while Pastor *et al*. (2008) have used sentence length and word length – calculated by readability indices – as a measure of simplification. In their study, the Automated Readability Index, the Coleman-Liau Index, and the Flesch-Kincaid Grade Level Test have been used. For the present study, the differences between the average reading scores of the three

subcorpora were used to assess the simplification hypothesis. The Flesch Reading Ease Test will be used for this purpose.

The Flesch Reading Ease Test is a component of Microsoft Word. It works by assigning a score on a scale that ranges from 0 to 100; the lower the score, the less readable the text is. The standard reading difficulty level ranges from 60 to 70, and texts with a higher score are considered easier to read while those with scores below this level are more difficult to read (Williams, 2005:153). The Flesch Reading Ease Test uses both sentence and word length (in syllables) for the reading difficulty calculation, where the formula used is: 206.835 – (1.015 x average sentence length) – (84.6 x average number of syllables per word).

An example of a text with a high readability score of 80 from the IT subcorpus reads:

> At the next corner a surprise was waiting for me: the dark coloured Volkswagen Kombi had turned at the previous street and drove around the block to meet me! A huge fellow was sitting behind the steering wheel. Marcellus Wallace, the one who caused havoc in Pulp Fiction, wasn't a patch on him. He had a huge clean shaven head, big knuckles on the steering wheel, and a thin stripe of glittering teeth in his dark face.

An extract of a text with a lower readability score of 38.5 that is taken from the ET subcorpus reads:

> Section 110c - [. . .] (2) Any person who, intentionally or through negligence, contravenes any provisions or prohibitions that may have been provided by law for the fulfilment of the state's obligations as a member of the United Nations shall be liable to a fine or to imprisonment for any term not exceeding four months or, in aggravating circumstances, to imprisonment for any term not exceeding four years.

3.5.2.3 Word length (S3 operationalisation)

Average word length, on its own, has also been used as an operationalisation to detect simplification in studies such as those of Yuan and Gao (2008) and Kruger and Van Rooy (2012). This is based on the assumption that texts with an average shorter word length are easier to understand than texts with a longer average word length, since word length generally correlates with morphological complexity. Average word length, measured in characters, was included as operationalisation for simplification in this study as triangulation for the S2 operationalisation. Average word length in characters for each

text was calculated by means of the WordList function of WordSmith Tools 5.0, and compared across the three subcorpora.

3.5.3 Normalisation

3.5.3.1 Contraction ratio (N1 operationalisation)

The use of full forms rather than contractions may, according to Olohan (2004:101), be investigated as an indication of both explicitation and normalisation. In the first instance, contractions along with *that*-omission discussed in Sectionn 3.5.1, constitute reduced surface forms. According to Biber *et al*. (1998:150) "a reduction in surface form also results in a more generalized, less explicit content". Olohan (2001:431) takes this as an indication that if surface forms in translations are not reduced to the same degree as surface forms in non-translations, translations are "more explicit, less generalised in both form and content", which therefore constitutes explicitation.

In terms of normalisation, contractions can be used as a measure of translators' supposed tendency to normalise to a perceived conventionalised standard. Contractions tend to be associated with spoken and informal registers, and are generally discouraged in more formal and written registers (Kruger & Van Rooy, 2012:38). The use of full forms, rather than contractions, is therefore regarded as a feature of the written standard. The use of full forms as opposed to contractions in translation may therefore be seen as a reflection of the tendency to normalise "to a conservative interpretation of the conventionalised written standard" (Kruger & Van Rooy, 2012:38).

Based on Olohan (2001), a list of contractions and their full forms was drawn up, distinguishing between verb contractions and *not*-negation contractions (see Table 3.13). The two groups of contractions were investigated separately in order to determine whether there are different preferences for the use of the full or contracted form depending on the type of contraction. The Concord function of WordSmith Tools 5.0 was then used to extract all the relevant instances from the three subcorpora. The occurrence of contracted forms as a ratio of all possible instances where a choice could have been made was analysed.

Verb contraction	Verb full form
could've	could have
he'll	he will
here's	here is
he's	he is / has
he'd	he had / would
I'll	I shall / will
I'm	I am
I've	I have
I'd	I would
it's	it is / was
it'd	it would
it'll	it will
let's	let us
she'll	she will
she's	she is / has
she'd	she had / would
that's	that is / has
they'd	they had / would
there's	there is / has
there'll	there will
they'll	they will
they're	they are
they've	they have
we'd	we would
we'll	we shall / will
we're	we are
we've	we have
what's	what is / was
who'd	who had / would

Not-negation contraction	Full form
aren't	are not
can't	cannot
couldn't	could not
didn't	did not
doesn't	does not
don't	do not
hadn't	had not
hasn't	has not
haven't	have not
isn't	is not
mustn't	must not
needn't	need not
shan't	shall not
shouldn't	should not
wasn't	was not
weren't	were not
won't	will not
wouldn't	would not

who’re	who are
who’s	who is / was
you’d	you had / would
you’ll	you will
you’re	you are
you’ve	you have

Table 3.13: List of contractions and full forms

3.5.3.2 Neologisms and loanwords (N2 operationalisation)

A number of studies (Kenny, 2001; Kruger & Van Rooy, 2012; Williams, 2005) have used the notion of lexical creativity as a means to detect normalisation. This approach is based on the tenet that since normalisation makes the language of target texts more conventional and less creative than that of non-translations, it may be expected that creative or otherwise “non-standard” lexical items will occur to a lower degree in translations. According to Kenny (2001:73) lexical creativity can be found in either the formation of new words or the innovative collocation of existing words. Investigating neologisms (also known as coinages), defined as “the creation of a new word out of existing elements” (Crystal, 2002:455), can therefore give a useful indication of lexical creativity. For the purposes of this study, neologisms were defined as new words whose forms are not recorded in standard, general dictionaries (Fischer, 1998:4).

In addition to neologisms, the frequency of loanwords may also be used as an indicator of normalisation. Loanwords are lexical elements which are borrowed from another language or dialect because they fill a particular gap in the recipient language, or because they, in some way, embody prestige (Shulka & Connor-Linton, 2006:492). Since loanwords are peripheral to the core vocabulary of a language, and since the excessive use of loanwords is often discouraged in the standard written form of a language, the frequency of loanwords may also be regarded as a measure of normalisation or conservatism.

In order to identify neologisms and loanwords for investigation, the methodology employed by Kruger and Van Rooy (2012) was used. As neologisms and loanwords will most likely occur infrequently, hapax legomena, or word forms which occur only once (Kenny, 2001:130), were used as the investigation set. The WordList function of WordSmith Tools 5.0 was used to generate a list of

all the different words in the corpus, together with their frequencies. The hapax legomena were then copied into Microsoft Word where the lexicalised entries identified by the program's spellchecker were deleted. After this, the following entries were removed: all spelling errors, abbreviations, attested compounds – whether hyphenated or not, acronyms, proper nouns, and parts of e-mails. The remaining words were checked on the Internet and Oxford Dictionaries Online for attested usage.[9] Oxford Dictionaries Online was chosen as it allows one to check usage in the British, American, and world context. The remaining entries were tagged as neologisms or as loanwords.

3.5.4 Levelling-out

3.5.4.1 Register variation (L1 operationalisation)

Linguistic features demonstrate differential distribution according to register, leading to what is termed "register variation", or "the distinctive ways in which linguistic features are relatively common or rare, when compared to the use of those features in other registers" (Biber *et al*., 1998:137). It may be hypothesised that if translations do tend to gravitate towards the centre of a continuum, as assumed by the proposed feature of levelling-out, the register variability that is a characteristic of non-translated language may be reduced in favour of a more neutral middle register. In the case of this study, the assumption is that if levelling-out does indeed occur, registers in the translational subcorpora will show less variation for the features investigated than in the NT subcorpus. In other words, this means that there will be a less strong register effect for the features investigated in the translational subcorpora than is the case for the NT subcorpus. If this should be the case, it may be assumed that translation is the cause of the levelling-out of the variation. For this operationalisation, the data for all the other operationalisations were used to determine whether there is less variability across the different registers in the translated subcorpora as compared to the NT subcorpus.

For each of the operationalisations outlined above, a hypothesis can be formulated regarding the expected frequency in the three subcorpora. These hypotheses predict the effect of non-translated versus translated status, as well as the level of translation experience, on the occurrence of the feature in question. The hypotheses formulated for the study are set out in the following section.

[9] For more information visit http://oxforddictionaries.com. (Date of access: 08 August 2012.)

3.6 Hypotheses

3.6.1 Explicitation

In the light of the previous chapter, it may be proposed that the two translation subcorpora will demonstrate a higher degree of explicitness than the NT subcorpus. One of the reasons why explicitation is expected to occur with a higher frequency in the translated subcorpora is as a consequence of the interpretation process performed by translators (Blum-Kulka, 1986:292), when the ideas conceived in one language have to be reformulated in another language (Klaudy, 2008:107). The word "process" indicates that explicitation may occur unconsciously as it refers to the "involuntary departure from the information given in the source text, evoked by the specific conditions of cognitive language processing during language mediation" (Baumgarten *et al*., 2008:181). Translations are then seen as bearing traces of the interpretation process since translators tend to write down markers for all that they have construed (Pym, 2005). The more difficult the source text is to understand, the harder translators have to work to make sense of it, which means that they are more likely to make their renditions and interpretations explicit (Pym, 2005). Since this interpretation process is performed by both inexperienced and experienced translators, it may be proposed that linguistic indicators associated with explicitation will occur to a higher degree in translations than in non-translations.

In addition, it is expected that there will be a difference in the degree to which explicitation occurs in translations produced by inexperienced translators and those of experienced translators. This difference may be attributable to the different contexts and constraints imposed on the two groups of translators. Experienced translators, for instance, have to be aware of the element of risk involved in their work (Chesterman, 2011:177). According to Pym (2005), risk can be seen as the "probability of undesired outcome" where undesired outcomes refer to factors which restrict the cooperation between the communication partners involved. This restricted cooperation might result in translators not getting paid or losing the client (Pym, 2005). Therefore, it may be expected that experienced translators will aim "to do their best to reduce to a minimum the damage related to the transfer process, and it seems better to end up with some redundancy than with major losses" (Øverås, 1998:576). Inexperienced translators, in contrast, are inclined to take more risks, because, amongst other reasons, they fear the potential loss of credibility less than experienced translators (Künzli,

2004:45), and therefore they will be less likely than experienced translators to make use of explicitation.

In addition to the possibility that explicitation is the consequence of risk management, it has also been postulated that explicitation might occur as a result of translators' assumptions about their roles as cultural and literary mediators, which are spurred by their suppositions about their readership and target audience (Saldanha, 2008:28). In this sense, it may be that translators resort to explicitation because they realise that their readership, unlike the source text readers, have fewer shared cultural references (Pym, 2005), or less common ground with the author, and may therefore need help with the interpretation of the text. Similarly, the occurrence of explicitation is also thought to be a consequence of the fact that translators, as both readers and writers, are aware of the difficulty of meaning construction (Pym, 2005). Hence, their concern with coherence and clarity may prompt them to provide missing links, lay implicit meanings bare, explain complicated passages and elaborate on the source text (Kruger, 2002:90) so as to promote the target audience's understanding of the text. In general, it may be expected that inexperienced translators will not be as sensitive to the needs of their target audience as experienced translators, simply because they lack the experience to appreciate the importance of their roles as cultural and literary mediators and the importance of taking their readership into account. Based on this, it may be proposed that their translations will contain explicitation to a lesser degree than those produced by experienced translators.

For the E1 operationalisation, it is hypothesised that there will not only be a higher frequency of the optional syntactic element that (and thus a lower omission ratio) in the translation subcorpora than in the NT subcorpus, but it will also be more frequent in the ET subcorpus than in the IT subcorpus. In other words, it is foreseen that the ET subcorpus will have the lowest omission ratio of the three subcorpora, and the NT subcorpus the highest. In the case of the E2 operationalisation, it is hypothesised that the three types of conjunction markers will be most prevalent in the ET subcorpus, less common in the IT subcorpus and least frequent in the NT subcorpus.

3.6.2 Simplification

Like explicitation, simplification is expected to occur to a higher degree in the two translated subcorpora than in the NT subcorpus. This supposition is based on the postulation that simplification is an inherent part of the cognitive activity involved in the language mediation process (Laviosa,

2002:46). This cognitive approach to simplification can be explained from within a cognitive grammar approach (Halverson, 2003). Based on this theory, translators are likely to restrict their lexical choices to a limited range of the available options because of an asymmetry present in their schematic networks (Halverson, 2003:222). This asymmetry occurs because some nodes are used more than others, which means that some links are more prominent than others (Halverson, 2003:222). Once established, these prototypes will be selected more often than those which have been used less frequently, and this will consequently result in simplification at the lexical level (Halverson, 2003:222). In other words, this line of thought leads to the notion that simplification will be noticeable to some degree in all translations since it is an inherent part of the translation process.

While simplification might be evident to a higher degree in all translations (as compared to non-translated texts), the experience of translators may have an impact on the frequency with, or degree to which it occurs. In this respect it is important to bear in mind that experienced translators, on the one hand, will most likely have more distinguished prototypes since their greater experience means they will have had plenty of opportunity to reinforce some nodes and links more than others. Inexperienced translators, on the other hand, are likely to have less established networks with weak links, and will thus lack clearly defined prototypes. Therefore, from a viewpoint based on cognitive grammar, it may be assumed that experienced translators will simplify more than inexperienced translators.

Logically, however, this theory has to be questioned, for the reason that experienced translators have more skill and will have developed more resources to draw on, including a greater vocabulary range. As discussed in Section 2.2.2.1, language competence is considered to be one of the components of translation expertise. This competence involves vocabulary, morphological, syntactic, and graphemic knowledge along with an understanding of linguistic variation – such as regionalisms or dialects – as well as knowledge of cultural reference and an awareness of figures of speech. Therefore it may be assumed that contrary to the prediction based on prototype theory, translations produced by experienced translators will demonstrate a larger vocabulary range and lexical diversity compared to those produced by inexperienced translators who have not yet had as much opportunity to develop the language component of translation expertise.

It is also important to consider that experienced translators will most likely have developed greater skill at extracting or reformulating the ideas expressed by the source text. Likewise, it is also possible that they cognitively simplify the message during the interpretation process to make sense of it and to understand it for themselves before they convert it for the target audience – a process which may be reflected in their translations. Inexperienced translators, on the contrary, have been known to concentrate on the "surface of the translation problem" or "on the lexical surface of the source text and not on its underlying message or communicative intent" which means the linguistic structures of the source text tend to be more evident in their translations (Shreve, 2002:164). Based on this, it may be assumed that the meaning of the underlying message will be more difficult to understand in their translations due to the influence of the source text. In this view the greater source-text interference characteristic of translations produced by inexperienced translators correlates with a lower degree of simplification.

As an additional result of the cognitive factors involved in the translation process, simplification may also be a consequence of the fact that translators wish to clarify the communicative intent of the source text (Blum-Kulka & Levenston, 1984:119) and consequently to make "things easier for the reader" (Baker, 1996:182). In order to do so, they have to contend with the fact that different languages often lack a comparable organisation of lexis along with comparable cultural concepts (Laviosa, 2002:46). These lexical and cultural mismatches force translators to resort to strategies in order to bridge the gaps between source and target languages (Laviosa, 2002:46), which often results in simplification. As experienced translators tend to be more aware of target text acceptability criteria than inexperienced translators (Fraser, 2000:112), it follows that they take more relevant aspects into account with the aim of ensuring that they produce translations that fulfil their function for the specified target readers (Göpferich & Jääskeläinen, 2009:174). The experienced translators' greater awareness will likely prompt them to compensate for mismatches between the source and target language which will result in simplification. As inexperienced translators are also less experienced in the assessment of translation, which is often evidenced by the fact that they mainly focus on semantic or lexical acceptability criteria (Frasier, 2000:112), and source text structures as explained above, it may be postulated that they will simplify to a lesser degree than experienced translators.

Based on the above, a hypothesis was formulated for each of the three simplification operationalisations. For the S1 operationalisation it is hypothesised that the IT subcorpus will have a

lower standardised TTR, and consequently less lexical diversity, than the ET subcorpus. The NT subcorpus will have the highest standardised TTR. In the case of the S2 operationalisation, it is predicted that the ET subcorpus will be the easiest to read, followed by the IT subcorpus, with the NT subcorpus the most difficult in comparison. Lastly, for the S3 operationalisation it is hypothesised that the texts of the ET subcorpus will have the shortest word lengths overall, while the NT subcorpus will have the longest word lengths with the IT subcorpus somewhere in between.

3.6.3 Normalisation

The occurrence of normalisation in relation to the level of translation expertise is difficult to predict, as the literature suggests a number of potentially conflicting hypotheses. In general, normalisation as a recurrent feature of translated language is believed to be the result of the socio-cultural or economic constraints which are applicable to translation (Kenny, 2001:67). Translators are routinely faced with the possibility that their translations may be rejected, criticised or ignored by their intended target audiences if their translations deviate from norms of acceptability in the target language and culture (Kenny, 2001:67). These notions of acceptability are also likely to be reinforced by editors and other publishing agents who are aware of the higher financial risk translations carry if they are not well received by the target audience, and therefore pressure translators to produce target texts that appear "normal" (Kenny, 2001:67). All of these factors may eventually cumulate in translators' conscious or unconscious internalisation of market preferences and target language norms which require fluent translations (Kenny, 2001:67). Based on the above, it may be hypothesised that normalisation to target language convention will be more evident in the work of experienced translators who regularly deal with these constraints and, as a result, more likely avoid taking risks.

Inexperienced translators, in contrast, do not experience the same constraints and, as mentioned previously, are more inclined to take risks, as they are less concerned with a loss of credibility if their interpretation is wrong. Furthermore, they do not yet have the same ability to distinguish between instances where risk-taking is inappropriate or should be avoided (Künzli, 2004:45). In continuation, Moe (2010) argues that normalisation, particularly in terms of register, may also be a consequence of lack of experience. According to Moe (2010:132), translations tend to "become more stylistically neutral, or more target oriented" which is often reflected in register changes which result in texts that resemble the neutral normative variants of non-translations produced in that particular target language. One of the possible reasons given for the production of a stylistically neutral target text is

the inexperience of the translator who produced it (Moe, 2010:132). Thus, a hypothesis for the occurrence of normalisation in the translations of inexperienced translators is complicated by the suggestion that they are expected to normalise less than experienced translators because they are more inclined to risk-taking, yet at the same time they may be more inclined to normalise as they are less sensitive to register requirements than experienced translators are.

The occurrence of normalisation, as is the case with simplification, has also been explained by approaches based on cognitive grammar. According to this approach, the gravitational pull between the category prototype and the highest-level schema might contribute to the tendency of translators to overrepresent or exaggerate the typical features of a target language (Halverson, 2003:221). This overrepresentation occurs because the specific target language grammatical and lexical structures correspond to the salient configurations and nodes in the schematic network (Halverson, 2003:218). However, in cases where the source language structures are weakly and distantly linked, if at all, to the target language structures, an underrepresentation of the typical target language features tends to occur (Halverson, 2003:223). Thus, according to this theory, normalisation may be more prevalent in the translations of experienced translators who are expected to experience a greater gravitational pull. Inexperienced translators, alternatively, are expected to have weaker links and connections, which will result in less normalisation.

Normalisation makes the language of translations "more conventional, less marked and less creative than the language of source texts and non-translated texts" (Castagnoli, 2008:22). Creativity, therefore, occurs to a lower degree as translators opt for more conventional and conservative target language solutions to problems created by unusual or creative source text features (Kenny, 2001:66). Yet, creativity is considered to be one of the components of translation expertise (PACTE, 2011) and, as such, is a feature that experienced translators can be expected to utilise when the need arises. Moreover, even though inexperienced translators are more inclined to take risks, instances of highly creative jumps are likely to alternate with run-of-the-mill solutions (Neubert, 2000:13). In sum, if less creativity is indicative of normalisation, it may be assumed that inexperienced translators will normalise to a greater degree than experienced translators who will be able to make use of creative solutions when necessary.

Clearly, the tensions between perspectives on risk-taking, cognitive aspects, and creativity suggest conflicting hypotheses about the relative prevalence of normalisation in the translations of experienced and inexperienced translators. For this study, it is hypothesised that experienced translators will normalise to a lesser degree than inexperienced translators as they are not only more creative, but are also expected to be more sensitive to the register of the source text and the specific function it serves, which is indicative of their inclination to preserve it in their target texts. The non-translated texts are expected to contain the lowest degree of normalisation since their authors most likely did not experience the same cognitive, socio-cultural or economic constraints as translators in the text-production process.

In terms of the N1 operationalisation, it is expected that contracted forms will occur less frequently in the IT subcorpus than in the ET subcorpus, as it is believed that experienced translators are more sensitive to register variability than inexperienced translators, and contraction use is strongly linked to register variation. Contractions are expected to occur to the highest degree in the NT subcorpus. For the N2 operationalisation, it is hypothesised that neologisms and loanwords, as markers of lexical creativity, will be more prevalent in the ET subcorpus than in the IT subcorpus, but will occur most frequently in the NT subcorpus.

3.6.4 Levelling-out

Predictions regarding the occurrence of levelling-out are more difficult to make than for the other features of translation discussed. The difficulty posed by levelling-out is the consequence of the fact that little empirical testing has been done to test the validity of the concept, and very few translation theorists have commented on it (Williams, 2005:45). Therefore, there is little evidence which supports the overall hypothesis (Olohan, 2004:100) or evidence which can be used to formulate a prediction of the frequency of its occurrence in this study.

Pym (2008:10) suggests that if translations do share the recurrent features of translation, in terms of explication, simplification and normalisation, it follows that translations will be inclined towards the centre of a continuum, which constitutes the fourth feature, levelling-out. In this sense, it might be speculated that levelling-out may be more prevalent in translations produced by experienced translators as it is expected that they will use explicitation and simplification to a higher degree than inexperienced translators. In addition, it may also be argued that less variance can be expected to

occur in this group because experienced translators are more closely aligned with a professional standard and, as such, will be more strongly and consistently subject to the norms and conventions of this standard. However, the experience they have acquired also encourages a greater awareness of other factors involved in translation, such as text type and register, which may result in a greater degree of variation as they adjust their translations in accordance with these factors in order to ensure their product will meet the expectations of the target audience.

As with the experienced translators, it is difficult to make a prediction about the degree to which translations produced by inexperienced translators will demonstrate levelling-out. It may be postulated that less variation will be evident in inexperienced translators' translations based on the fact that they are expected to be less aware of the variety of factors that affect translation choices. In this sense, their translations may be expected to level out as they take the same type of approach to the translation task by mainly focusing on source-text structures rather than making allowance for differences such as register or text type variation. Conversely, it may be argued that because inexperienced translators have less exposure to and are consequently less aligned with the conventions of professional standards their translations will display a greater degree of variation.

Against the above background, it is tentatively postulated that levelling-out will occur to a greater degree in the work of inexperienced translators than experienced translators. For the L1 operationalisation, it is therefore proposed that the registers in the IT subcorpus will show the least variation for all of the features investigated. It is hypothesised that the ET will demonstrate more register-related variation whereas the NT subcorpus will have the highest degree of variation.

Once all of the subcorpora had been compiled and the various hypotheses made for the different operationalisations, it was necessary to analyse the texts to not only see the differences in the distribution and frequency of the various features, but also to determine whether these differences were statistically significant or not.

3.7 Data analysis

The data for each operationalisation (or dependent variable) in each text were captured in an Excel spreadsheet, together with information on the independent variables of corpus and register. All frequencies and type-token ratio (TTR) were normalised to 1000 words.

For each operationalisation, descriptive as well as inferential statistics are reported. For each analysis, it is first determined whether the data have a normal distribution, using a chi-square test. If the data are normally distributed, descriptive statistics are reported using means (as a measure of central tendency) and standard deviations (as a measure of dispersion). In instances where data are normally distributed, a parametric test to determine the statistical significance of the differences between the three subcorpora may be used. In this study, one-way ANOVA is used for this purpose. For each feature, it is indicated whether the results are statistically significant or not, and the nature of the difference between the three subcorpora is explained by reporting differences in means.

If the values for an operationalisation do not have a normal distribution, descriptive statistics are reported in the form of a boxplot illustrating median values (as a measure of central tendency) and interquartile range (as a measure of dispersion). Since normal distribution of values is an assumption of most parametric tests, significance testing for operationalisations demonstrating non-normal distribution of data has to be done by means of a non-parametric test. An ANOVA is unsuitable because ANOVA assumes that the different groups or experimental conditions' variances are homogeneous and are uncorrelated with the means (Hill & Lewicki, 2006:229). In this study, the Kruskal-Wallis ANOVA (a non-parametric equivalent of ANOVA) is used for significance testing in cases where data are not normally distributed. It will be noted if the results are statistically significant before the nature of the differences between the three subcorpora are discussed in terms of differences between mean ranks. The Kruskal-Wallis ANOVA works by ranking the entire data set before the ranks are compared across the different groups (Oxford University Press, 1997:351). As such, the mean ranks indicate which samples have lower or higher scores. The main purpose of mean ranks, as a measurement of central tendency, is to identify the value which would be most typical of a full set or in this case of the specific feature tested for each of the subcorpora (Oxford University Press, 1997:351).

There is one exception to this general procedure: the L1 operationalisation, which uses register variation as an index of levelling-out, investigates the interaction between the independent variables of corpus and register, for each of the dependent variables used in the other analyses. In instances where data are normally distributed, a factorial ANOVA is used for assessing whether there is a statistically significant interaction between the variables corpus and register (which indicates that the

three subcorpora demonstrate different register-related preferences for a particular feature). If there is a statistically significant interaction effect, the data are interpreted to determine whether the translation subcorpora demonstrate less register variation. Features where data are not normally distributed ordinarily require a non-parametric test for significance; however, there is no non-parametric test to test whether interactions between variables exist. In cases like these, an appropriate power transformation has to be applied to the dependent variable to stabilise the variance and to eliminate or reduce the correlation between the standard deviation and the means (Hill & Lewicki, 2006:229). In this study, where possible, the data were transformed using a Box-Cox transformation to normalise the distribution, after which a factorial ANOVA was performed on the normalised data. In cases where data transformation was not possible, a factorial ANOVA was performed. The results are adequately reliable, since there were sufficient data points for each level of the factorial ANOVA (more than 10, generally), and since p-values were either very large or very small.

In each of the analyses, the aim of the statistical test is to determine if there are statistically significant differences between the subcorpora. In other words, for each operationalisation it is determined if the null hypothesis of no difference between groups can be rejected with confidence, using a confidence level of $p<0.05$.

3.8 Conclusion

In this chapter, an overview was provided of the research design which, as a corpus-based study, contains both quantitative and qualitative elements. As corpus composition is an important concern, attention was given to the compilation of the corpus, specifically focusing on how the texts are comparable to one another in terms of registers, word count and source languages, and how the three subcorpora are representative of translations produced by inexperienced and experienced translators, and of non-translations. Data collection was discussed before the operationalisations chosen for the study were outlined in more detail and the expected frequencies for the various operationalisations discussed as hypotheses. Finally, the statistical processing of the data for the various operationalisations used to test the hypotheses was discussed.

CHAPTER 4: FINDINGS AND DISCUSSION

4.1 Introduction

The central hypothesis of this study is that marked differences will be apparent between translations done by inexperienced and experienced translators, and also that evident differences will be observable between translated and non-translated texts. These differences, it is held, will be reflected in the frequency and dispersion of the features of translated language, such as explicitation, simplification, normalisation and levelling-out. It is assumed that these features may be considered as a reflection of the different translation strategies used by experienced and inexperienced translators.

In the first part of the chapter, the findings for each of the operationalisations chosen to study instances of explicitation, simplification, normalisation and levelling-out are discussed. For each operationalisation it is determined whether the null hypothesis of no difference between groups can be rejected with confidence, or alternatively that there is not enough evidence to reject the null hypothesis. In the second part of the chapter, these findings are discussed with reference to the hypotheses formulated for the different operationalisations (see Section 3.6).

4.2 Findings

4.2.1 *That*-omission (E1 operationalisation)

The omission or inclusion of the complementiser *that* was studied based on the premise that its inclusion can be seen as a linguistic indicator of explicitation, since it serves to make grammatical relations more explicit. A range of reporting verbs were used as search terms for the investigation to determine whether the translated subcorpora explicitated more frequently than the non-translated subcorpus, and whether the ET subcorpus contained a higher degree of explicitation than the IT subcorpus, according to the hypothesis formulated for the E1 operationalisation.

In what follows, data are reported in the form of omitted instances as a ratio of all instances where a choice between the full or reduced form could have been made. As can be seen in Figure 4.1, the values for the *that*-omission ratio were skewed to the right because of the high number of 0-values, and the distribution of the data is therefore not normal.

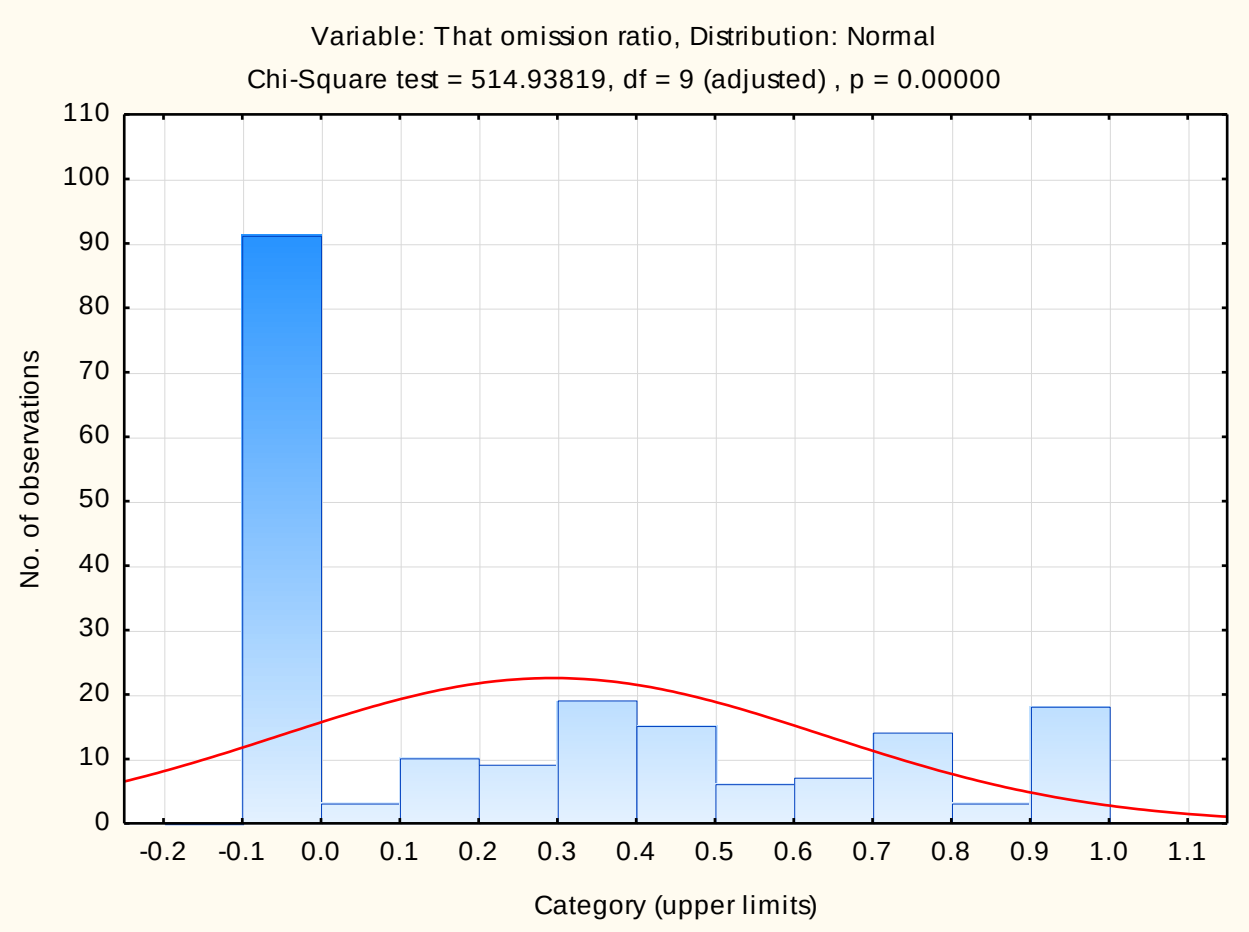

Figure 4.1: Non-normal distribution of values for the variable *that*-omission ratio

The boxplot in Figure 4.2 shows that the median value for *that*-omission for the IT subcorpus is 0.0, with the median value for the ET subcorpus slightly higher at 0.17. In the NT subcorpus the median value is higher, at 0.33. In other words, the NT subcorpus has a higher median than both the translated subcorpora while the IT subcorpus has the lowest median of the three subcorpora. These medians indicate that half of the documents in the IT subcorpus never omit the *that* complementiser, half of the ET subcorpus omit that less than 17% of the time, whereas half of the documents in the NT subcorpus omit the optional *that* less than 33% of the time. The upper quartile hinges of the three subcorpora indicate that three-quarters of the texts in the IT subcorpus omit the *that* complementiser up to 29% of the time while three-quarters of the ET subcorpus omit it up to 41% of the time. The upper quartile hinge of the NT subcorpus is the highest of the three, with three-quarters of the texts omitting it up to 73% of the time. These interquartile ranges for the three subcorpora indicate a greater degree of dispersion in the NT than the translated subcorpora, with the smallest degree of dispersion in the IT subcorpus.

Figure 4.2: Boxplot of *that*-omission ratio for the three subcorpora

To test whether these differences between the subcorpora are significant, a Kruskal-Wallis ANOVA was performed, which shows that the optional syntactic element *that* has been omitted most frequently in the NT subcorpus (with a mean rank of 111.9) whereas in the IT subcorpus it is omitted least frequently (with a mean rank of 82.8). The ET subcorpus has a mean rank of 95.5. The differences between the three subcorpora are statistically significant ($H(2, N=195)=9.60$, $p=0.008$).

These results support the E1 operationalisation hypothesis that the omission of the *that* complementiser is more prevalent in the NT subcorpus than in the two translation subcorpora, but not that the ET subcorpus would omit it more than the IT subcorpus. This means, in other words, that the NT subcorpus explicitated the least frequently of the three subcorpora, followed by the ET subcorpus, while the IT subcorpus explicitated the most frequently. Therefore, the hypothesis made for the E1 operationalisation is not entirely supported as even though the translation subcorpora explicitated

more than the non-translated subcorpus, the experienced translators explicitated less overall than the inexperienced translators.

4.2.2 Conjunctive markers (E2 operationalisation)

Conjunctive markers were used as an operationalisation of explicitation, since they mark instances of explicitation at the discourse level – guiding readers to see the relationships between propositions in the text. It was hypothesised that conjunctive markers would be most prevalent in the ET subcorpus, occur less in the IT subcorpus and the least in the NT subcorpus. The three types of conjunctive markers studied were markers of elaboration, extension and enhancement.

4.2.2.1 Elaboration

Since the values for conjunctive markers indicating elaboration were not distributed normally, but were skewed to the right (see Figure 4.3), a Kruskal-Wallis ANOVA was performed to test for significant differences between the three subcorpora. From this analysis it is clear that differences between the subcorpora for this variable are very small, and not statistically significant ($H(2, N=254)=2.11$, $p=0.35$).

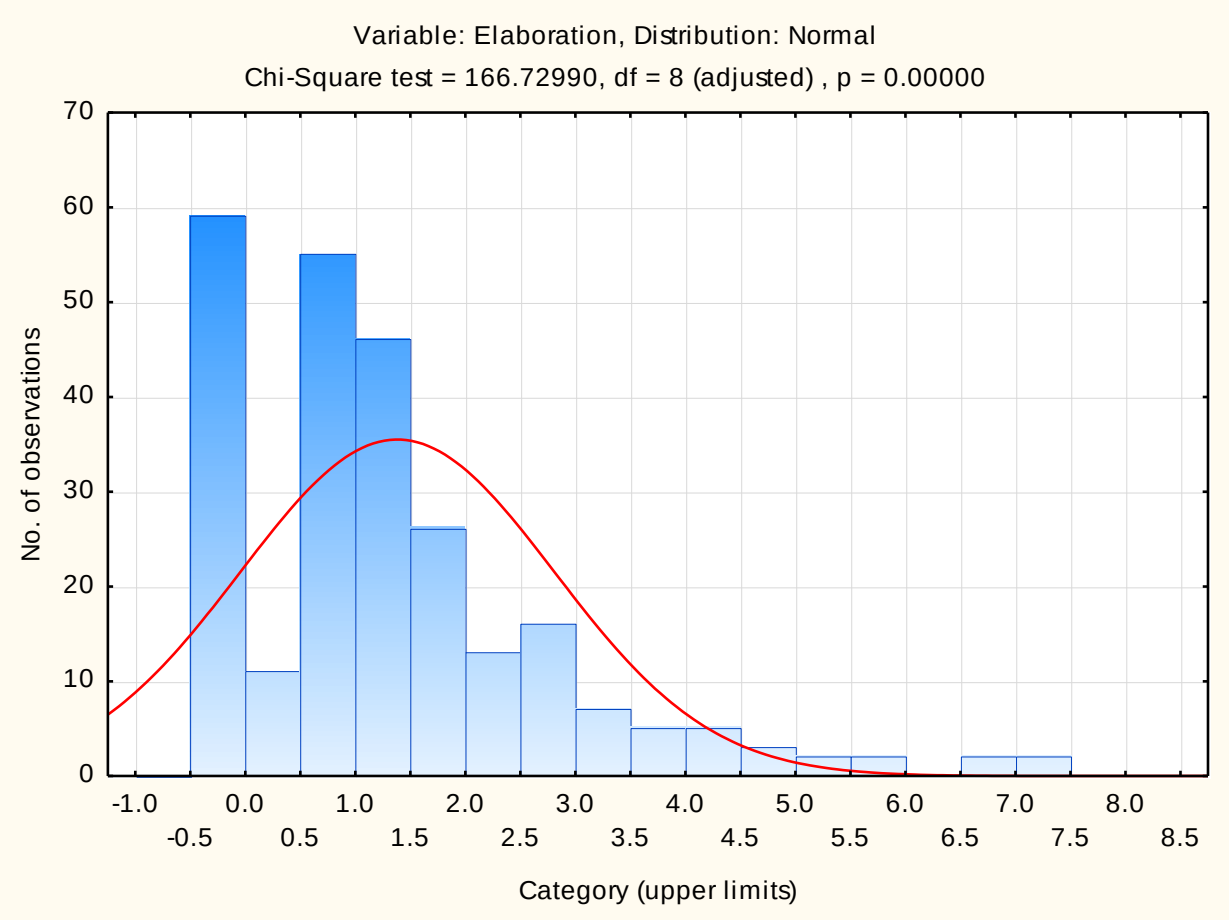

Figure 4.3: Non-normal distribution of values for the variable conjunctive markers indicating elaboration

The boxplot in Figure 4.4 demonstrates the similarity between the median values across the three subcorpora, with elaboration markers occurring at a median frequency of around 1 per 1000 words in all three subcorpora.

Figure 4.4: Boxplot of elaboration markers for the three subcorpora

4.2.2.2 Extension

The values for conjunction markers of extension were normally distributed (see Figure 4.5), and therefore a one-way ANOVA was used to test for significant differences between the three subcorpora.

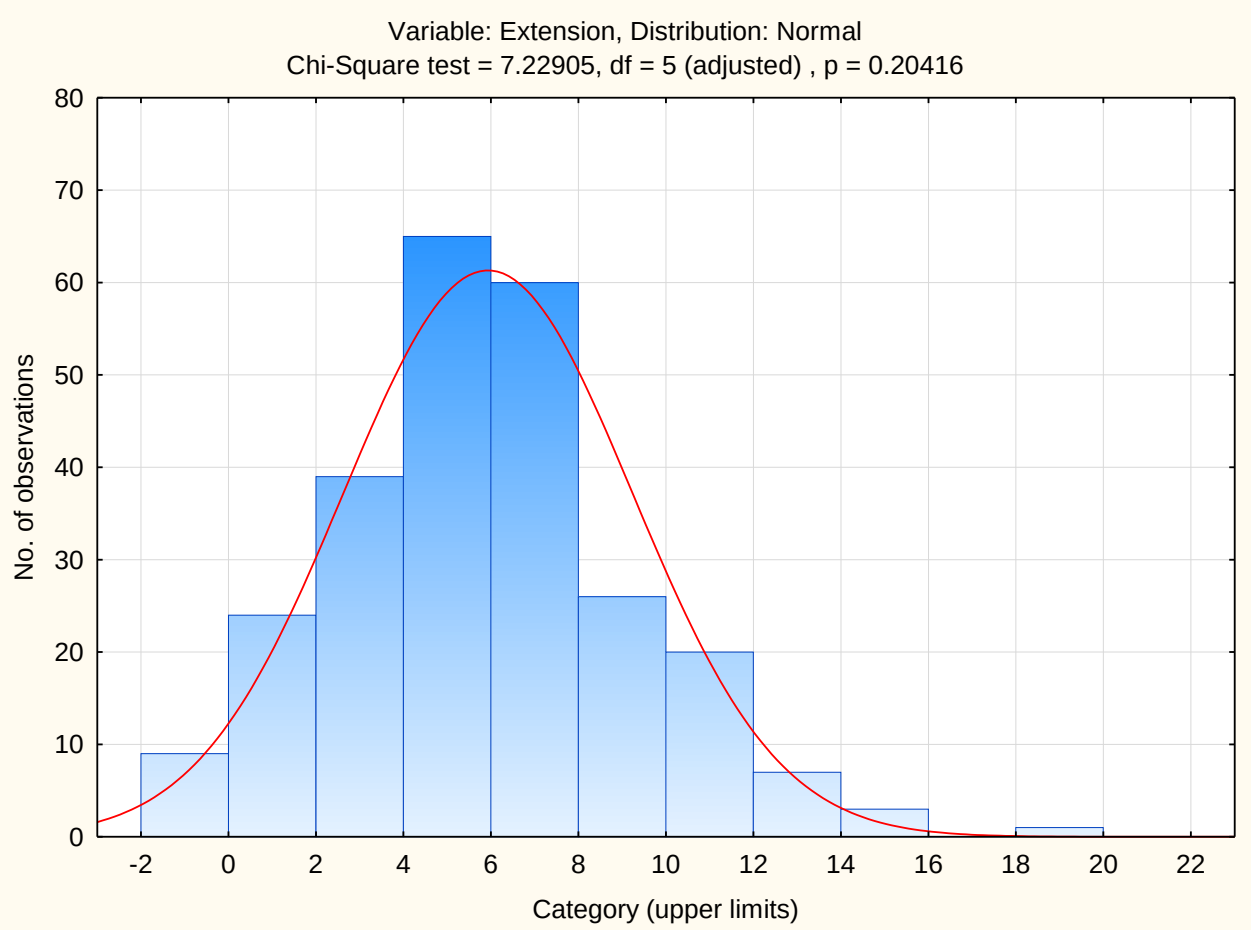

Figure 4.5: Normal distribution of values for the variable conjunction markers indicating extension

Using an ANOVA, the differences between the three subcorpora are not statistically significant ($F(2, 251)=0.04$, $p=0.96$), and therefore the null hypothesis (H_0) of no difference cannot be rejected with confidence. Figure 4.6 shows that the mean values for the three subcorpora are very similar to one another, with the NT subcorpus using extension markers only slightly more frequently than the two translation subcorpora, with mean values of around 6 per 1000 words in all three subcorpora.

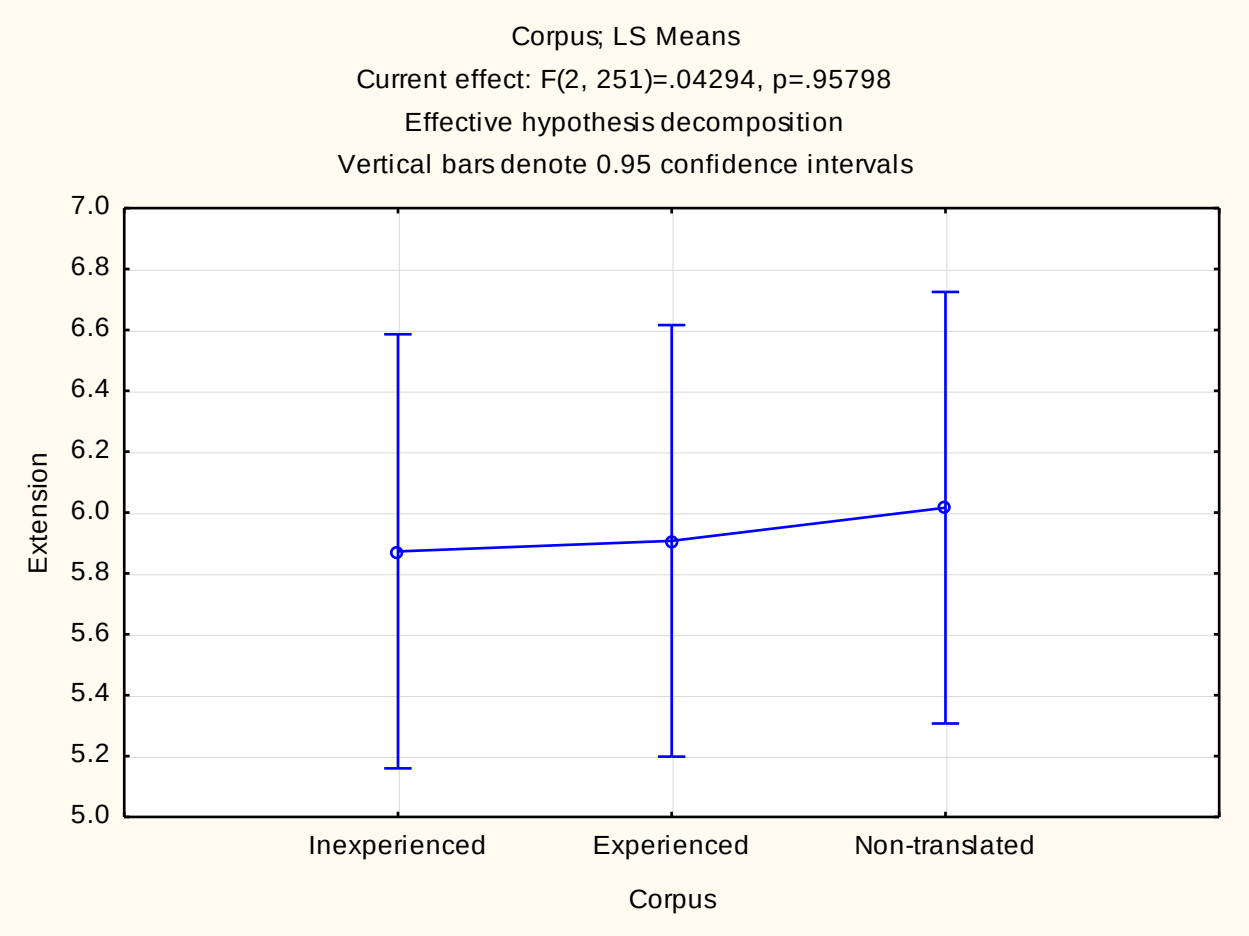

Figure 4.6: One-way ANOVA for conjunctive markers indicating extension in the three subcorpora

4.2.2.3 Enhancement

For the enhancement category of conjunctive markers, which consists of the causal-conditional, manner, matter and spatio-temporal sub-types, the values were not normally distributed, as depicted in Figure 4.7. A Kruskal-Wallis ANOVA was therefore performed to test for significant differences between the three subcorpora.

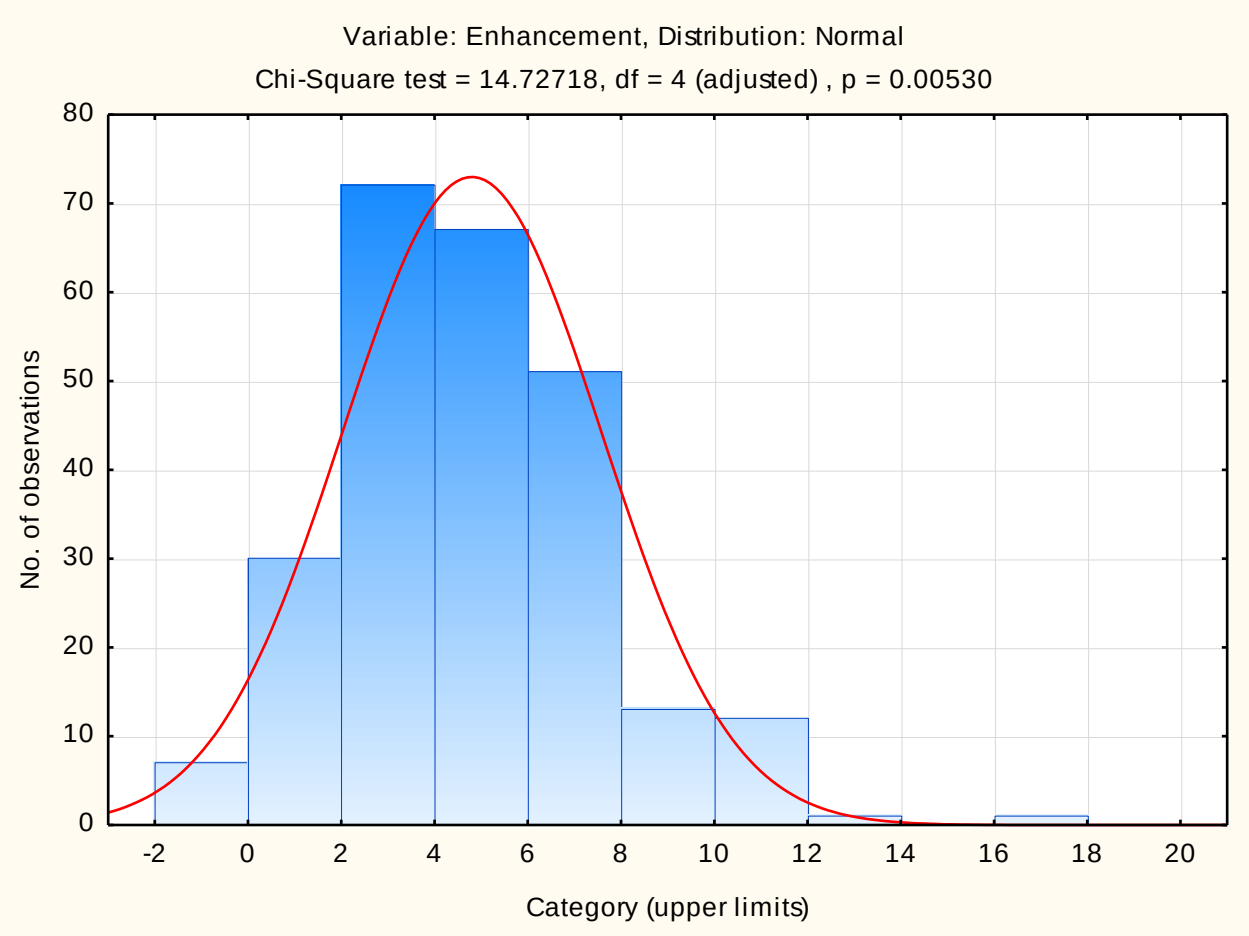

Figure 4.7: Non-normal distribution of values for the variable conjunctive markers indicating enhancement

The Kruskal-Wallis ANOVA shows that the differences between the three subcorpora are not statistically significant ($H(2, N=254)=0.88$, $p=0.64$). The boxplot in Figure 4.8 demonstrates the similarity of median values for this feature across the three subcorpora, with conjunctive markers of enhancement occurring at a median frequency of 4.4, 4.7 and 4.9 in the three subcorpora respectively.

Figure 4.8: Boxplot of enhancement markers in the three subcorpora

4.2.2.4 All categories of conjunctive markers

Since the distribution of values for all categories of conjunctive markers combined demonstrated a normal distribution (see Figure 4.9), a one-way ANOVA was used to test for significant differences between the three subcorpora.

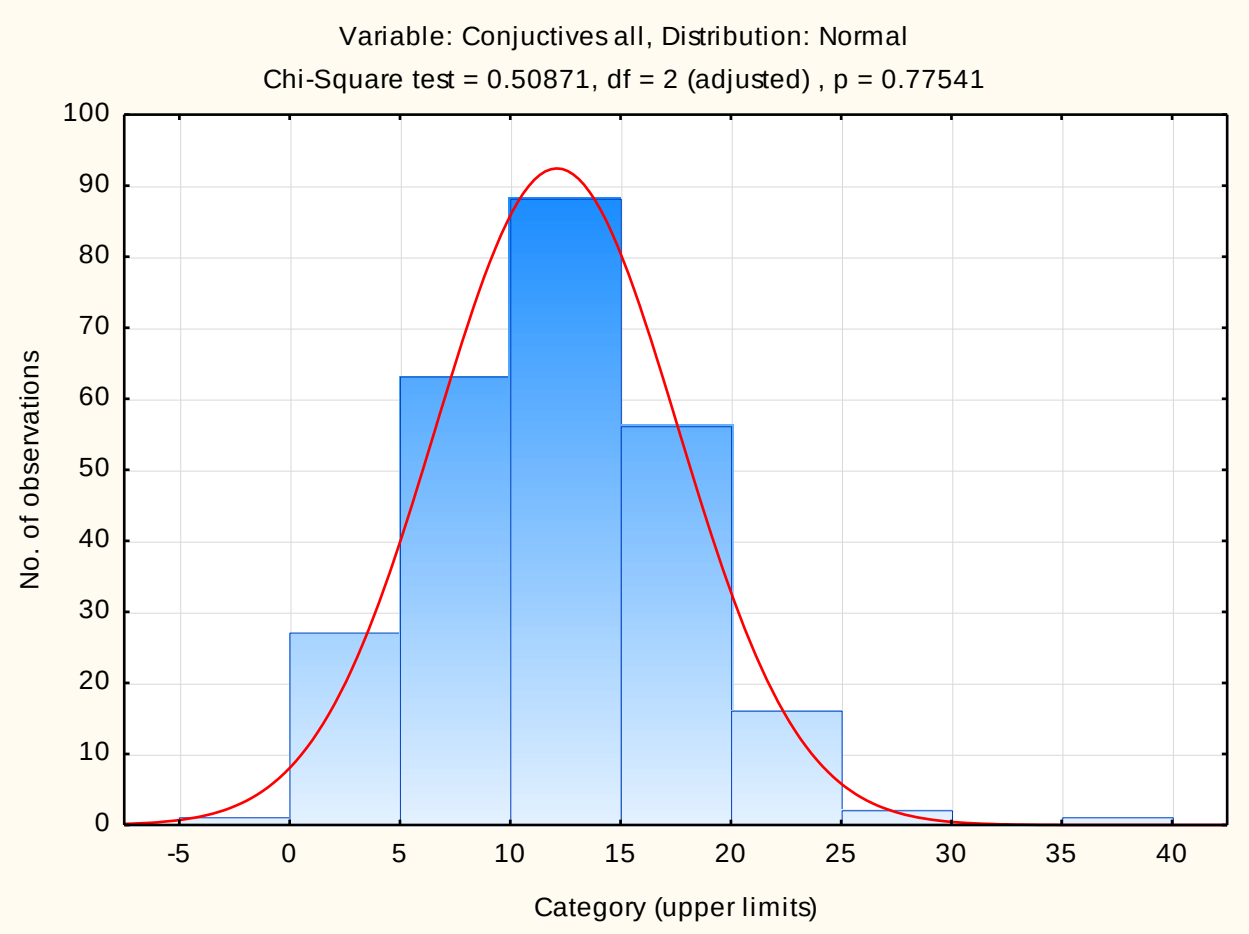

Figure 4.9: Normal distribution of values for the variable all conjunctive markers

Conjunctive markers were used least frequently in the IT subcorpus, at a mean frequency of 11.9 per 1000 words. The difference between means for the IT and ET subcorpora is negligible, with conjunctive markers occurring at a mean frequency of 12.0 per 1000 words in the ET subcorpus (see Figure 4.10). Conjunctive markers occur most frequently in the NT subcorpus, at a mean of 12.5 per 1000 words. However, the differences are small and not statistically significant ($F(2, 251)=0.26$, $p=0.77$) which means the null hypothesis (H_0) of no difference cannot be rejected. This means that there is no significant difference in the degree to which original text producers, experienced translators and inexperienced translators use explicitation at the propositional level.

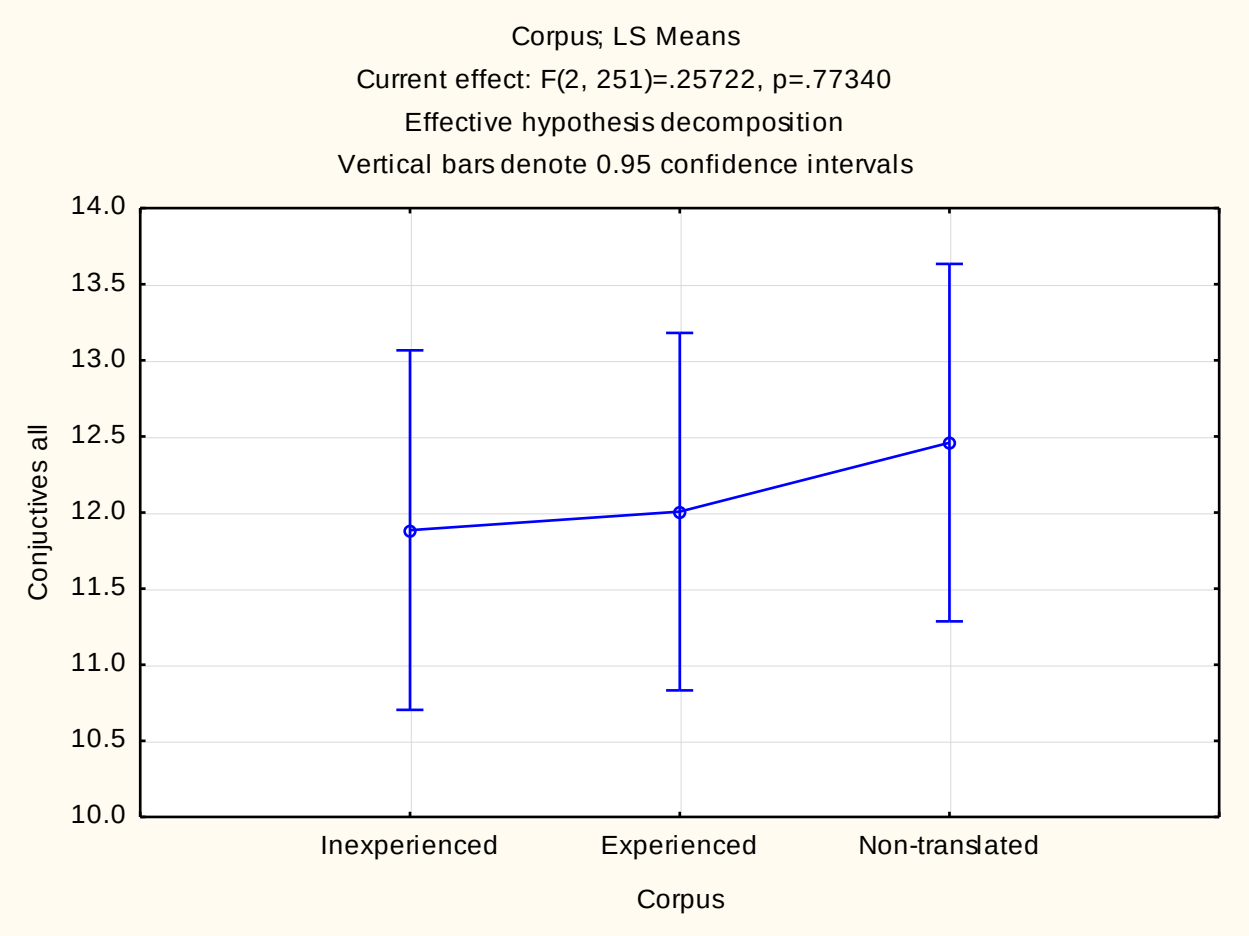

Figure 4.10: One-way ANOVA for all conjunctive markers in the three subcorpora

4.2.3 Type-token ratio (S1 operationalisation)

Standardised type-token ratio (TTR) was chosen as an operationalisation of simplification, since it measures the superficial lexical complexity of texts by focusing on their lexical diversity. A text with a lower standardised TTR is therefore less complex, as such a text has a smaller lexical range and is therefore easier to understand, at the lexical level, than one with a higher standardised TTR. For the S1 operationalisation it was hypothesised that the NT subcorpus would have the highest standardised TTR of the three subcorpora, followed by the ET subcorpus. The IT subcorpus was expected to have the lowest standardised TTR.

The values for the variable standardised TTR showed a non-normal distribution, as the values were skewed to the left, as can be seen in Figure 4.11.

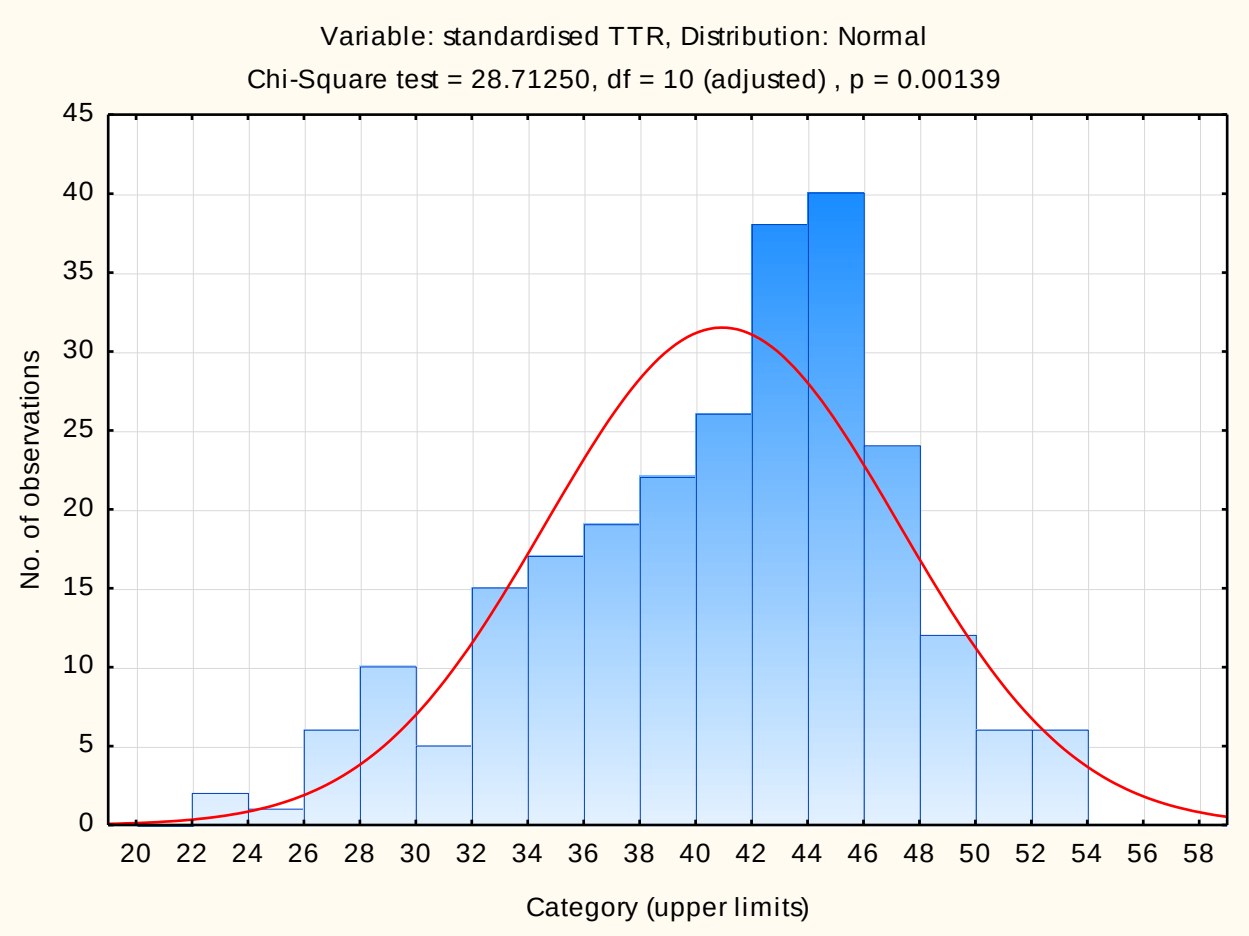

Figure 4.11: Non-normal distribution of values for the variable standardised TTR

The boxplot in Figure 4.12 indicates that half of the documents in the IT subcorpus contain at least 38% unique lexical items per 1000 words, which is the lowest of the three subcorpora, whereas the NT subcorpus has the highest standardised TTR of the corpus, with at least 45% unique lexical items occurring per 1000 words, in half of the texts. The median of the ET subcorpus shows that it is somewhere in between, as half of the texts contain at least 43% unique lexical items per 1000 words. In general, a somewhat greater degree of dispersion of the data can be noted in the translated subcorpora, with slightly larger interquartile ranges.

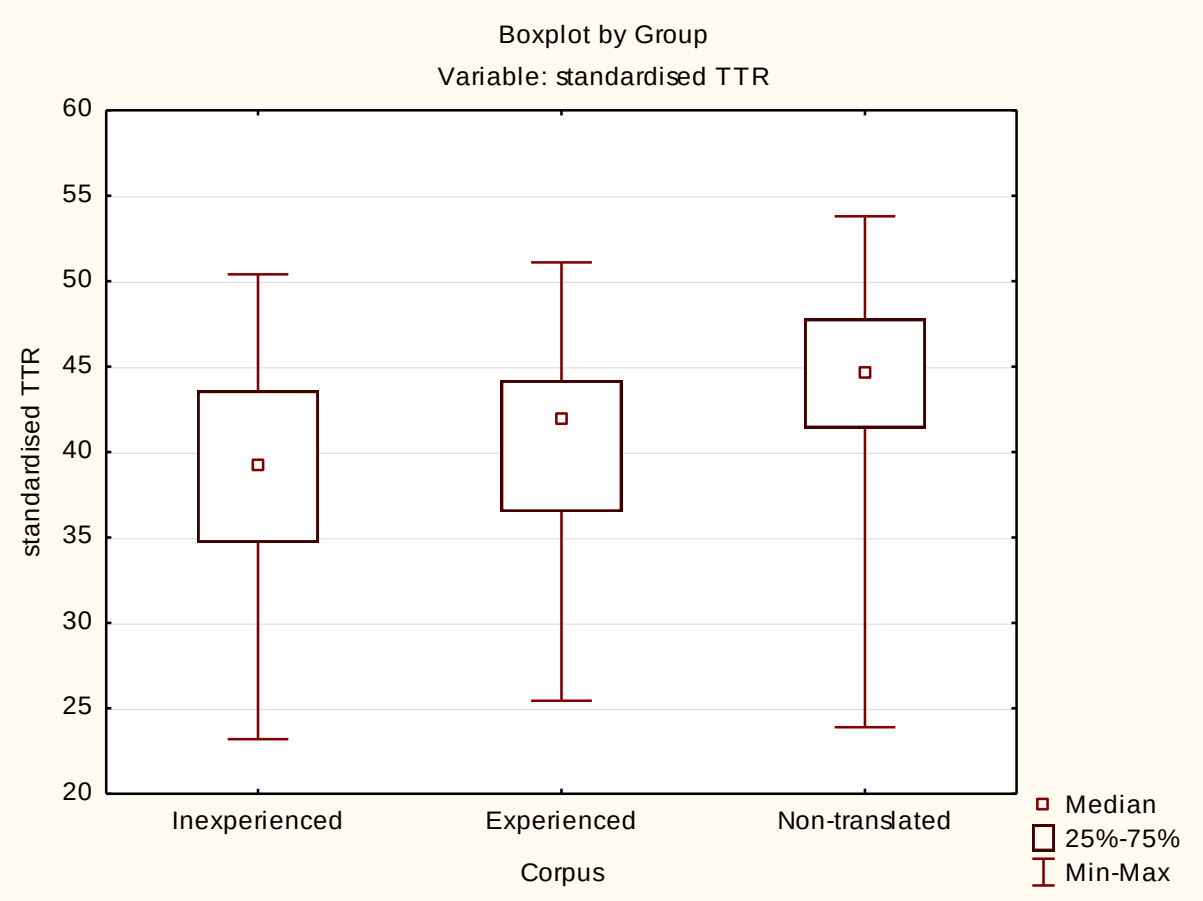

Figure 4.12: Boxplot of standardised TTR in the three subcorpora

The results of the Kruskal-Wallis ANOVA, which was done as the values for standardised TTR showed a non-normal distribution, show that the differences between the three subcorpora are significant ($H(2, N=249)=29.88$, $p<0.001$), which means the null hypothesis (H_0) of no difference can be rejected with confidence. Overall, the IT subcorpus had the lowest mean rank of the three subcorpora, at 97.4, and the NT subcorpus the highest, at 157.7. The ET subcorpus has a mean rank of 118.0. In other words, the NT subcorpus has the greatest range of vocabulary and lexical diversity of the three subcorpora, which means that the vocabulary is generally more varied in this particular subcorpus, compared to the two translated subcorpora. This greater lexical range also means that the non-translated texts in this particular corpus are more complex and more difficult to understand in comparison to the translated texts. The IT subcorpus, with the lowest score, is the easiest to read and to understand as it has the lowest lexical diversity when compared to the TTRs of the ET and IT subcorpora. As explained in Section 3.6.2, it was expected that the IT subcorpus would have a lower score than the ET subcorpus because inexperienced translators will not have had as much opportunity

as experienced translators to develop the language or linguistic competence of translation expertise. This competence involves, amongst others, vocabulary, syntactic, graphemic and morphological knowledge. There is therefore support for the S1 operationalisation hypothesis as it was expected that the IT subcorpus would have a lower standardised TTR, and consequently less lexical diversity, than the ET subcorpus, and for the NT subcorpus to have the highest standardised TTR.

4.2.4 Readability score (S2 operationalisation)

The readability score generated by the Flesch Reading Ease Test was used as operationalisation for the feature of simplification since it measures the reading difficulty of a text and/or the degree of skill required to read it. The readability score may therefore be used as an indication of the relative complexity or difficulty of a text, as measured by the various factors used to calculate the readability score (e.g. word length and sentence length). Since it is assumed that simplification is a feature of translated language, the translated subcorpora were expected to have higher scores (indicating greater ease of reading, and therefore greater simplification) as indicated by the Flesch Reading Ease Test than the NT subcorpus, with the IT subcorpus demonstrating a lower score than the ET subcorpus. In other words, it was expected that the NT subcorpus would demonstrate the lowest degree of readability (and hence the greatest complexity), and the ET subcorpus the highest (and thus the greatest degree of simplification). It was foreseen that the readability (and therefore complexity) of the IT subcorpus would fall somewhere in between that of the other two subcorpora.

As can be seen in Figure 4.13, values for the Flesch Reading Ease score showed a normal distribution.

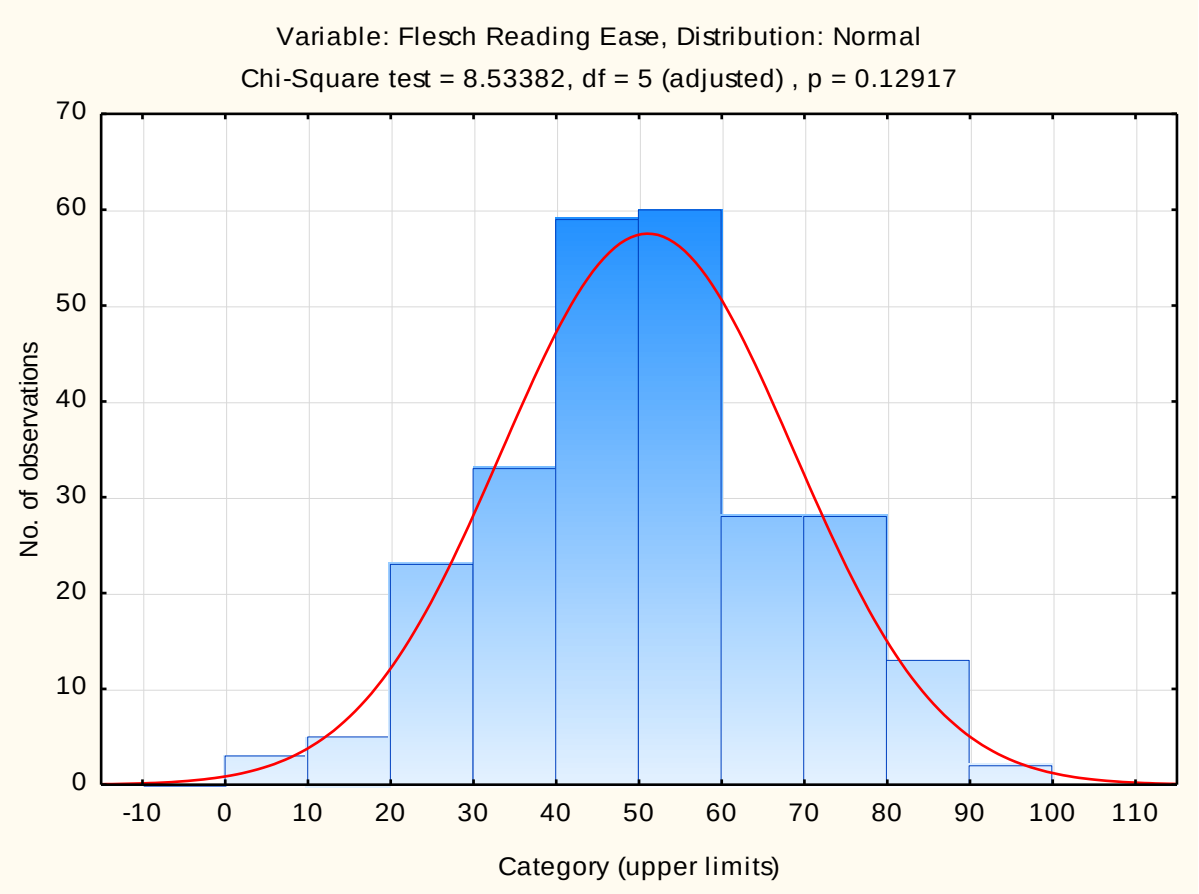

Figure 4.13: Normal distribution of values for variable Flesch Reading Ease score

The results of the one-way ANOVA (see Figure 4.14) show that the mean readability score of the ET subcorpus (47.40) was lower than that of the other two subcorpora, with the IT subcorpus demonstrating the highest mean value (53.25) followed closely by the NT subcorpus (52.59). This essentially means that the texts contained in the IT subcorpus are the easiest to read, as indicated by the highest mean readability score for this subcorpus. The most difficult texts to read are those in the ET subcorpus, with the lowest mean readability score. The difficulty of the NT subcorpus is only somewhat higher than that of the IT subcorpus, with the two subcorpora demonstrating very similar readability scores. The difference between the three subcorpora almost reaches statistical significance ($F(2, 251)=2.84$, $p=0.06$) which means that the null hypothesis of no difference can be rejected with some confidence. However, these findings do not support the S2 operationalisation, where it was expected that the ET subcorpus would be the easiest to read, followed by the IT subcorpus, and the NT subcorpus the most difficult in comparison. The findings for this operationalisation demonstrate that it is the ET subcorpus that is more difficult than both the IT and the NT subcorpora.

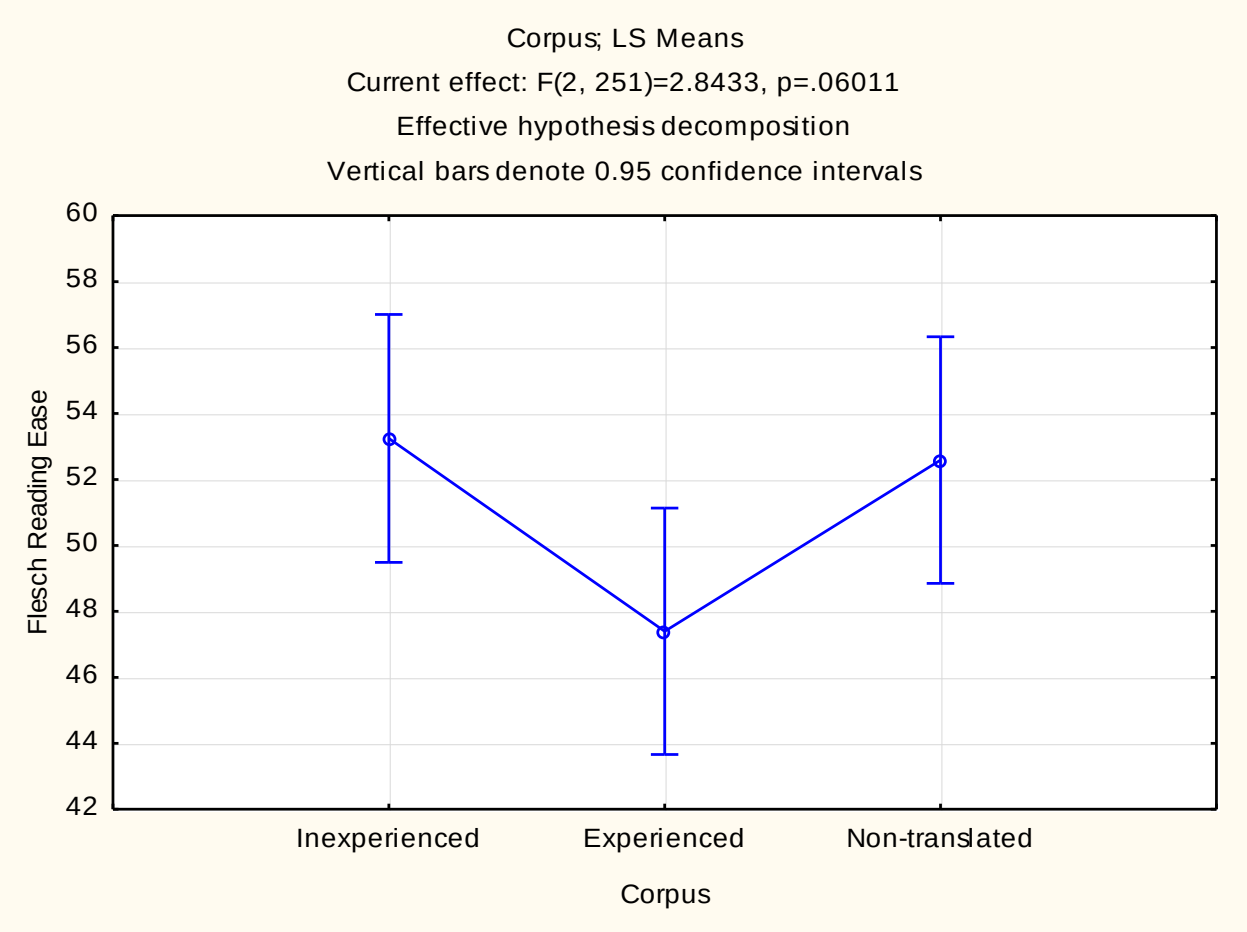

Figure 4.14: ANOVA for Flesch Reading Ease score in the three subcorpora

4.2.5 Word length (S3 operationalisation)

As the results of the S2 operationalisation were close to being statistically significant, the results of the S3 operationalisation, word length (measured in characters), were regarded as being of particular interest in validating the findings for the S2 operationalisation, since word length (measured in syllables) is one of the factors taken into account by the Flesch Reading Ease Test. Word length is taken as an indicator of simplification since it is believed that a shorter overall word length means the text is simpler and consequently easier to understand. It was hypothesised that the texts of the ET subcorpus would have the shortest word lengths overall, while the NT subcorpus would have the longest word lengths with the IT subcorpus somewhere in between.

The values for word length were normally distributed, as evident from Figure 4.15, and an ANOVA could therefore be performed.

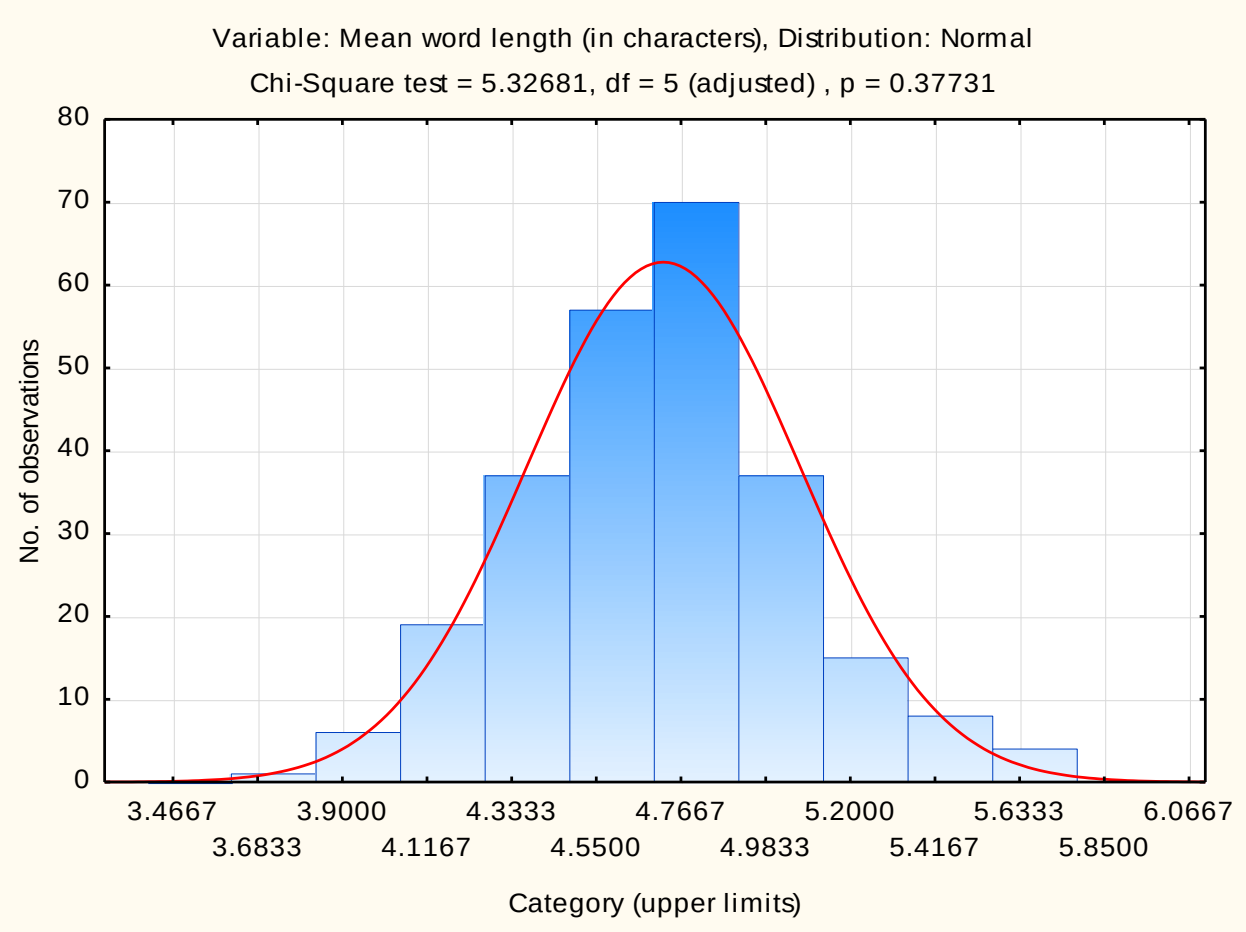

Figure 4.15: Normal distribution of values for the variable word length

The results of the ANOVA (see Figure 4.16) show that there is a significant main effect for corpus ($F(2, 251)=4.79$, $p<0.05$). However, the hypothesis formulated for the S3 operationalisation is not supported by the findings because, contrary to expectation, the mean word length of the ET subcorpus, at 4.81 characters, was higher than the mean word length of both the NT and IT subcorpora. Of the three subcorpora, the IT subcorpus had the shortest word length overall, with an average of 4.65, while the NT subcorpus was somewhere in between with a mean of 4.70. These findings support the findings of the S2 operationalisation. The IT subcorpus, with its shorter average word length also has the highest average readability score, and is therefore the least complex of the three subcorpora. The ET subcorpus has the longest words, on average, and also the lowest

readability score, and is therefore the most complex of the three subcorpora. The NT subcorpus, contrary to expectation, has word lengths and readability scores that place its complexity between that of the two translated subcorpora. Thus, the null hypothesis (H_0) of no difference cannot be rejected because the differences are not as predicted, significant as they may be. As is the case with the previous operationalisation, it appears that in terms of complexity, translations produced by experienced translators tend to be more complex than both texts produced by original text producers and translations produced by inexperienced translators.

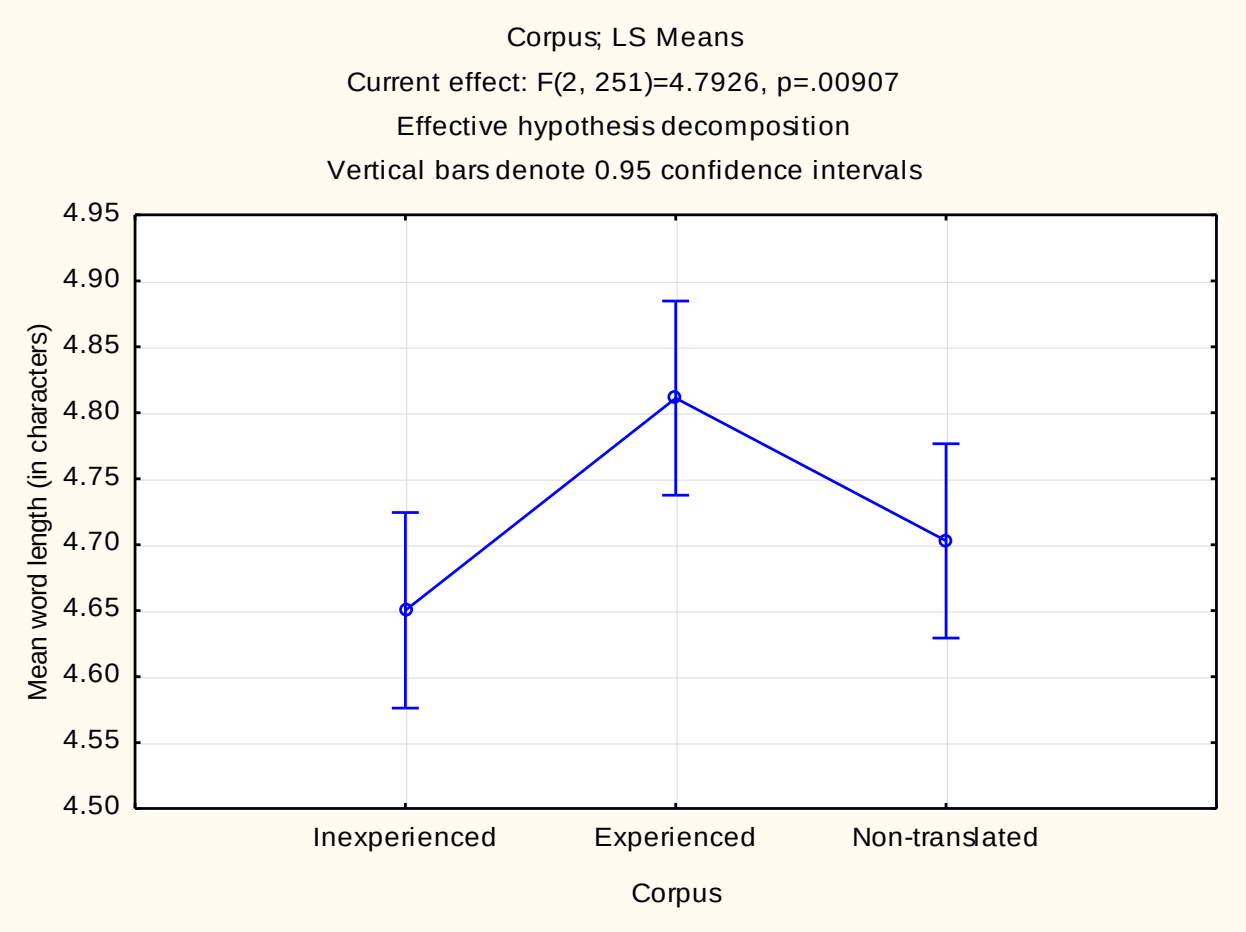

Figure 4.16: ANOVA for mean word length in the three subcorpora

4.2.6 Contraction ratio (N1 operationalisation)

The use of full forms rather than contracted forms was included as a measure of normalisation. Contracted forms are generally associated with spoken and/or informal registers, and are discouraged in many written registers. If translators are prone to normalise to the standard for written language,

they may be more inclined to avoid contracted forms, opting for full forms instead. This, in turn, would raise the formality of the translated text. For the N1 operationalisation it was hypothesised that contractions would occur most frequently in the NT subcorpus, less frequently in the ET subcorpus and least in the IT subcorpus, where the full form would mostly be preferred.

The contracted forms studied were divided into two groups, namely verb contractions and *not*-negation contractions.

4.2.6.1 Verb contractions

The values for the verb contractions, calculated as a ratio of all possible instances where a choice between a contracted or full form could have been made, were skewed to the right, as is evident in Figure 4.17.

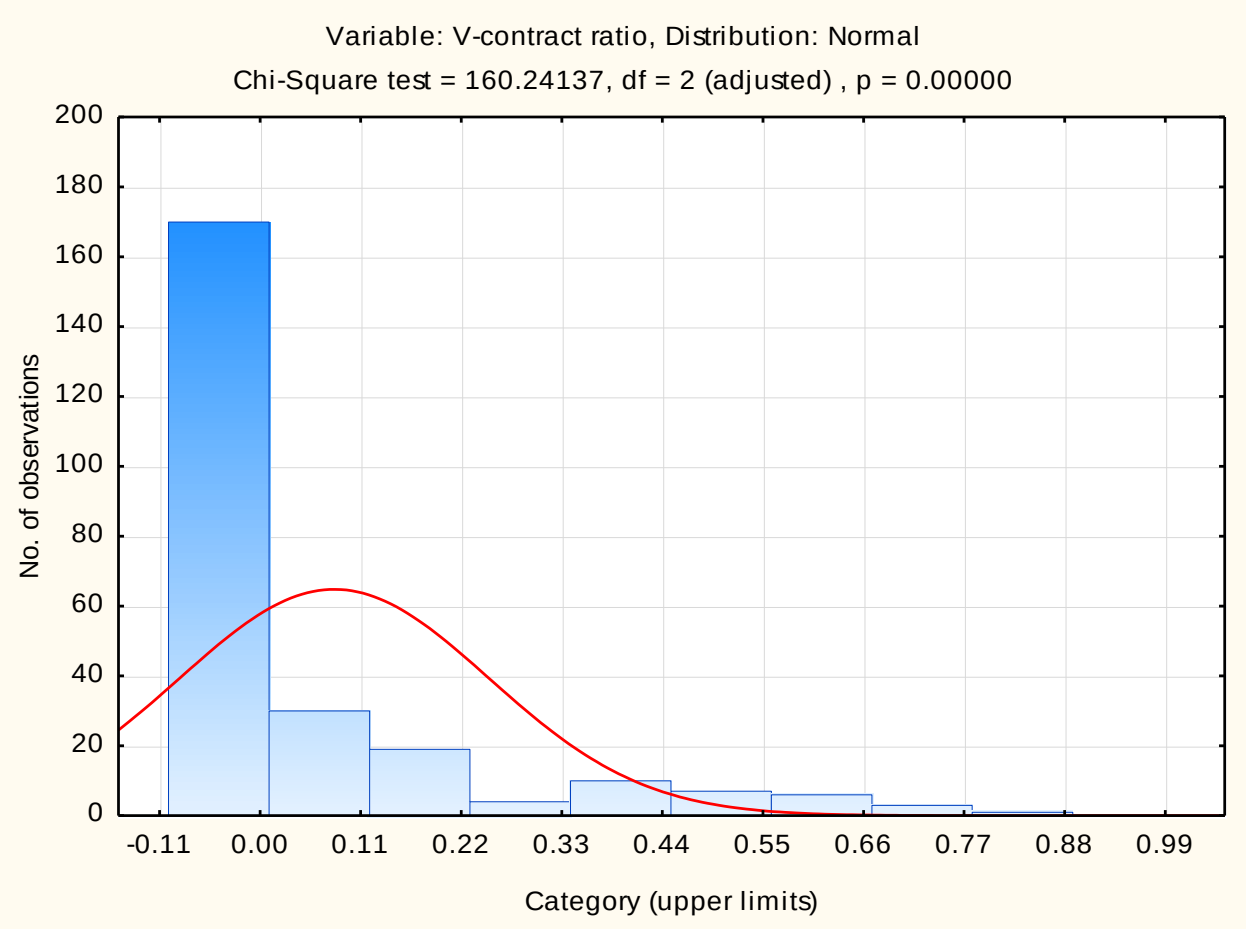

Figure 4.17: Non-normal distribution of values for the variable verb contraction ratio

The boxplot (Figure 4.18) shows that for the verb contraction ratio, the NT subcorpus has the highest upper quartile hinge (0.19), in comparison to the translated subcorpora, and its whisker also indicates the highest maximum data value in the three subcorpora (0.85). This indicates that there is a greater degree of variability in the NT subcorpus than in the other two subcorpora. Its median is also marginally higher at 0.01 than the medians of the translated subcorpora, which both rest on 0.0. In practice, this means that half the texts in this subcorpus use contractions up to 1% of the time, with three-quarters of the texts using it up to 20% of the time. As the translated subcorpora do not have upper quartile hinges and because their medians are situated on 0.0, it is evident that three-quarters of the translations in both subcorpora do not use verb contractions at all.

Figure 4.18: Boxplot of verb contraction ratio in the three subcorpora

A Kruskal-Wallis ANOVA shows that there is a significant main effect for corpus ($H(2, N=250)=20.68$, $p<0.001$). Overall, the translation subcorpora are less inclined to use verb contractions, with the IT subcorpus demonstrating a mean rank of 110.70, and the ET subcorpus of

115.93. The NT subcorpus uses contractions considerably more in comparison, with a mean rank of 149.52. This supports the hypothesis that the texts in the IT subcorpus would demonstrate the lowest frequency of contractions of the three subcorpora, the NT subcorpus the highest, with the ET subcorpus in between. From Figure 4.18, it can be seen that that the two translated corpora are quite similar to each other, with the NT subcorpus demonstrating the greatest difference.

4.2.6.2 *Not*-negation contractions

As with the verb contractions, values for the *not*-negation contraction ratio also demonstrated a non-normal distribution (see Figure 4.19).

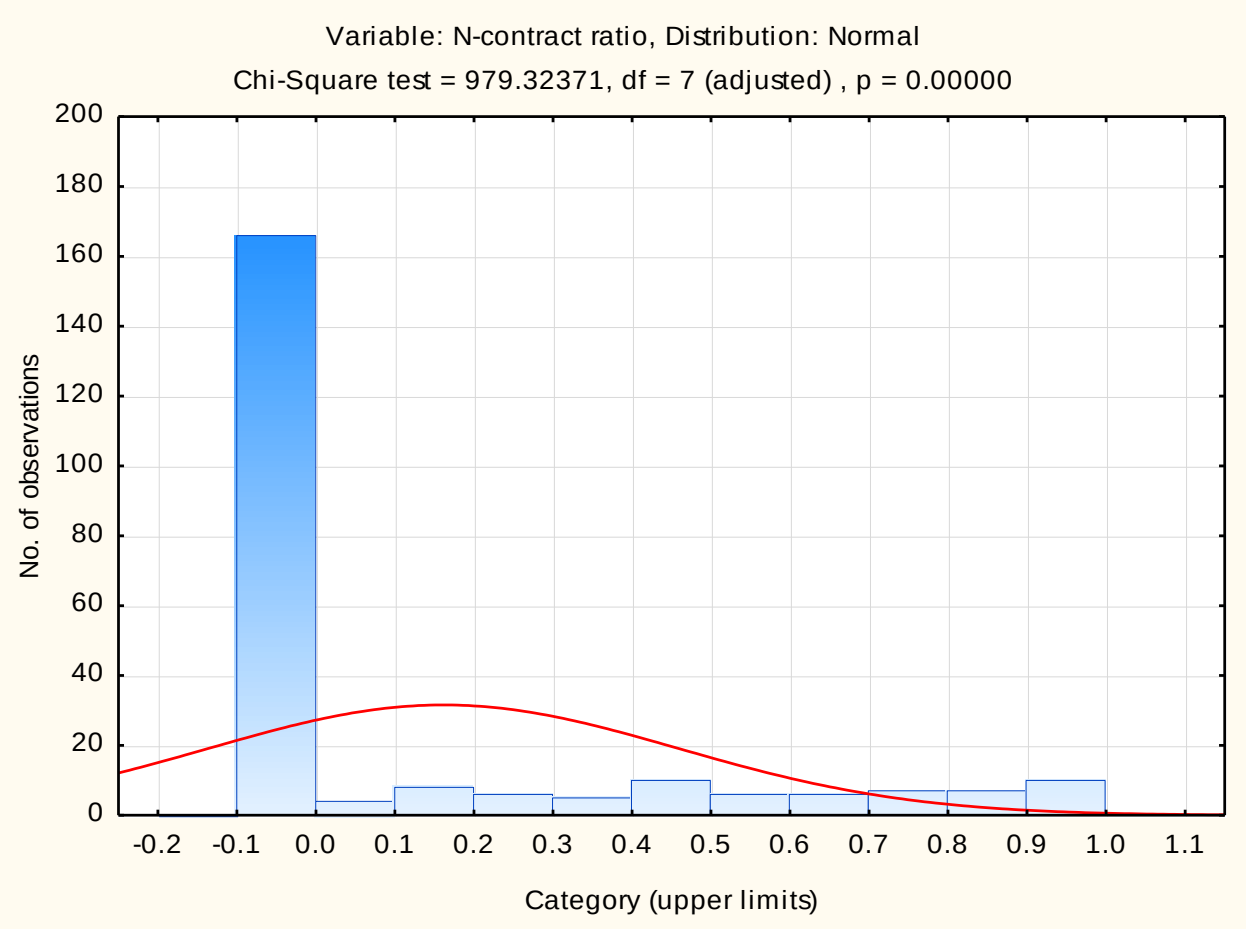

Figure 4.19: Non-normal distribution of values for the variable *not*-negation contraction ratio

Figure 4.20 indicates that for all three subcorpora the median is located on the 0.0 mark, which means that half of the texts in all three subcorpora do not use *not*-negation contractions at all. In addition, it

can be seen that the ET subcorpus has an upper quartile hinge of 0.1 whereas the upper quartile hinge of the NT subcorpus is noticeably higher at 0.52. The IT subcorpus, in turn, does not have an upper quartile hinge. These results indicate that three-quarters of the texts in the ET subcorpus use *not*-negation contractions up to 10% of the time. The NT subcorpus has a greater degree of dispersion, with three-quarters of the texts using it up to 52% of the time. As the IT subcorpus does not have an upper quartile hinge and because its median is situated on 0.0 it means that three-quarters of the documents contained in this particular subcorpus do not use contractions at all.

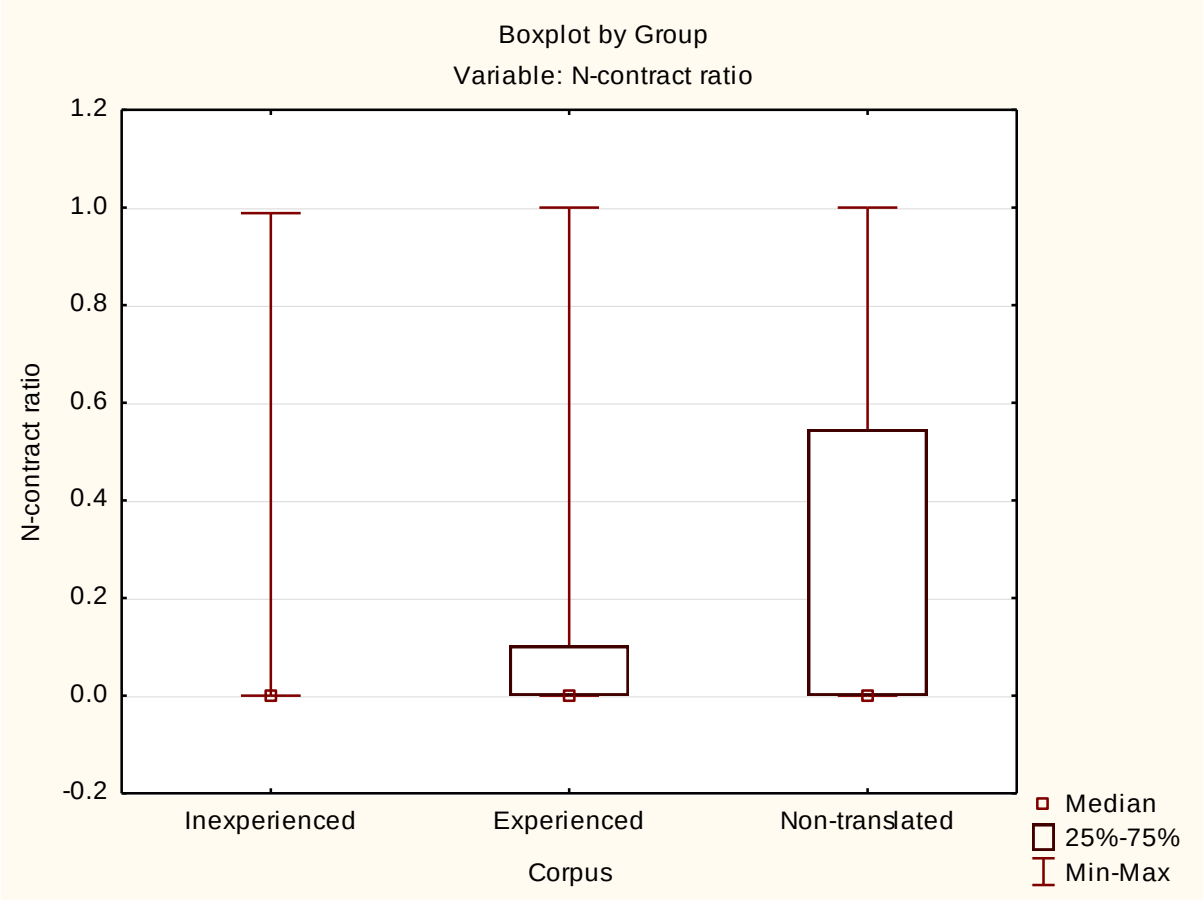

Figure 4.20: Boxplot of *not*-negation ratio in the three subcorpora

The Kruskal-Wallis ANOVA for *not*-negation contraction ratio shows a significant main effect for corpus ($H(2, N=235)=20.95$, $p<0.001$). Generally, the NT subcorpus (with a mean rank of 139.88) used contractions noticeably more than both the ET (mean rank of 113.74) and IT (mean rank of 100.60) subcorpora. The ranks of the two translated subcorpora are similar to one another, with the ET subcorpus using *not*-negation contractions slightly more than the IT subcorpus. This reflects the same order of frequency for verb contractions across the three subcorpora, and also reiterates that, for

this operationalisation, the most notable difference is between the translated subcorpora and the NT subcorpus.

4.2.6.3 All contracted forms

As the values for all the contracted forms combined were skewed to the right (Figure 4.21), a Kruskal-Wallis ANOVA was used to test for significant differences between the three groups.

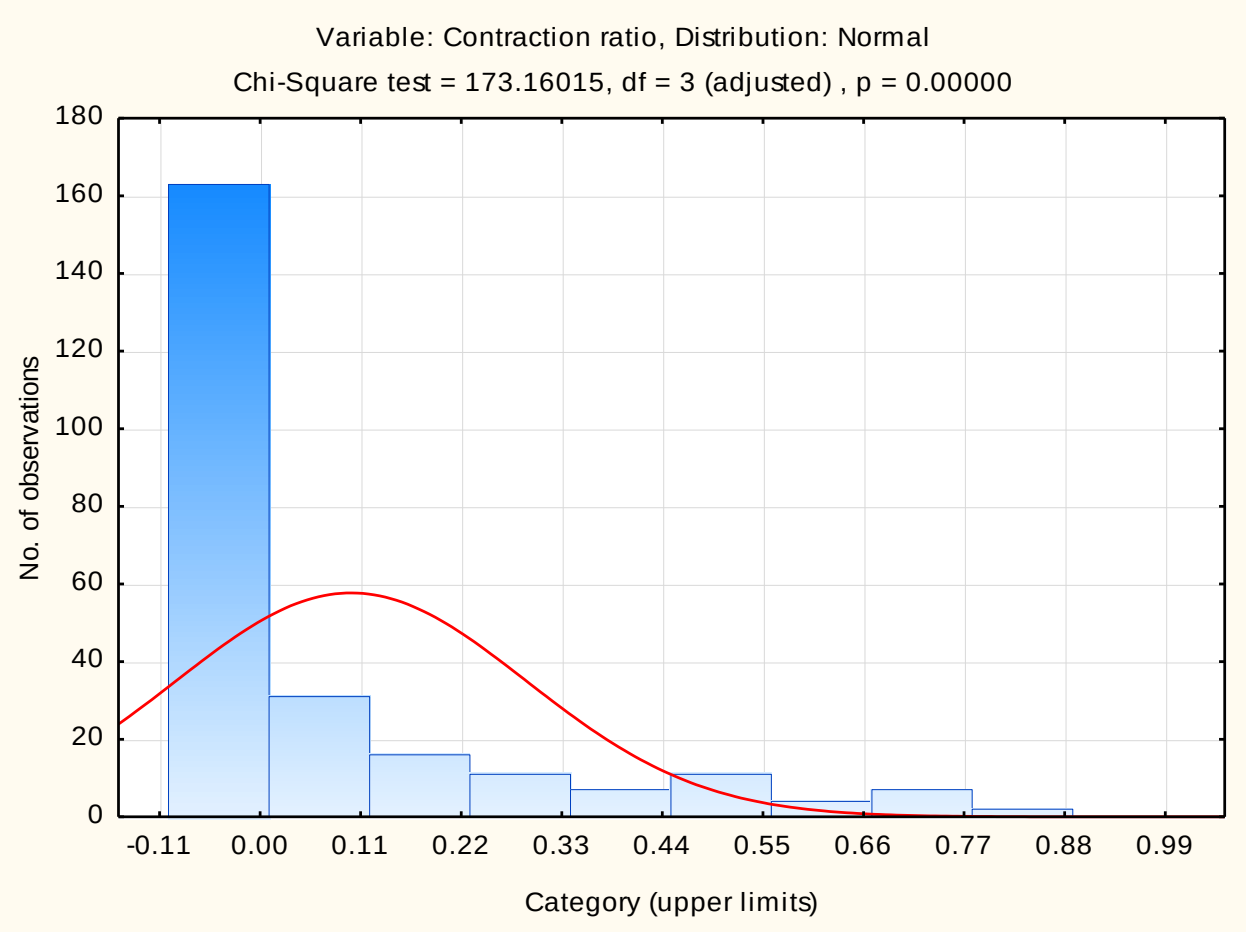

Figure 4.21: Non-normal distribution of values for variable contraction ratio

Figure 4.22 shows that when the values for the two categories of contracted forms are combined to arrive at an overall contraction ratio, the medians of the translated subcorpora are 0.0 whereas the median of the NT subcorpus is only slightly higher at 0.02. This means, in practice, that half of the texts in the two translated subcorpora avoid contracted forms altogether where a choice exists,

whereas in the NT subcorpus frequency is somewhat higher, with half of the documents in the NT subcorpus using contractions more than 2% of the time. The upper quartile hinge of the NT subcorpus is the highest of the three subcorpora, located at 0.28, followed by the ET subcorpus whose upper quartile hinge is 0.05, while the IT subcorpus does not have an upper quartile hinge. As the median of the IT subcorpus is located on 0.0 and because it does not have an upper quartile hinge it can be assumed that contracted forms are not selected for use in three-quarters of its documents. In the ET subcorpus three-quarters of texts use contractions up to 5% of the time. In the NT subcorpus, there is a bigger dispersion, with three-quarters of texts using contractions up to 28% of the time.

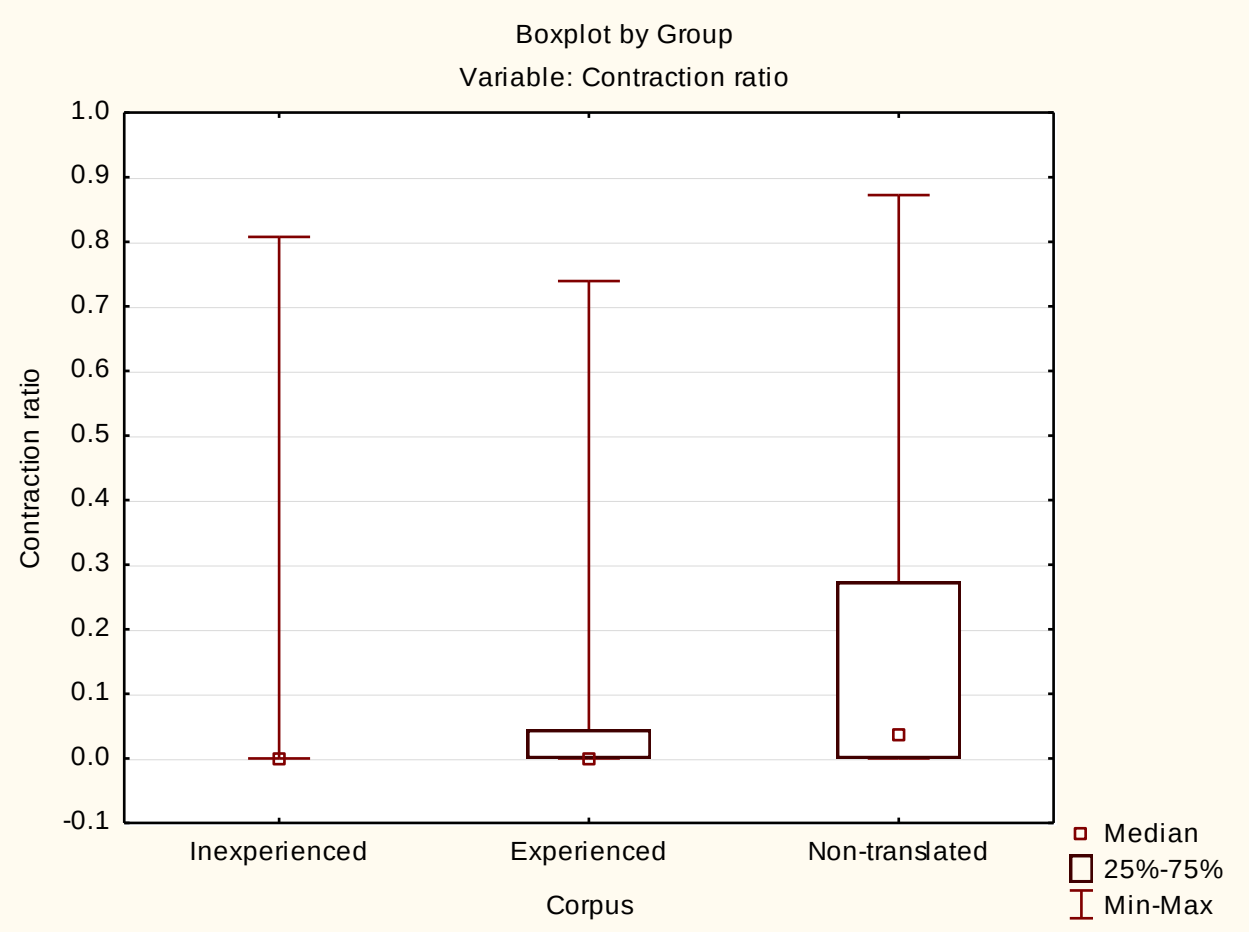

Figure 4.22: Boxplot for overall contraction ratio

A Kruskal-Wallis ANOVA for the overall contraction ratio shows a significant main effect for corpus ($H(2, N=252)=21.75$, $p<0.001$). The null hypothesis (H_0) of no difference, in other words, can be rejected with confidence. Overall, the NT subcorpus used contractions the most (with a mean rank of

151.98) whereas both the translation subcorpora used them significantly less. However, even though the scores of the translated subcorpora are similar to one another, the ET subcorpus used contractions slightly more frequently (with a mean rank of 117.55) than the IT subcorpus (with a mean rank of 109.88). The significant findings for both categories of contractions, as well as the two categories combined, can be taken as support for the hypothesis formulated for the N1 operationalisation, since the NT subcorpus shows less avoidance of contracted forms when compared to the translated subcorpora, and contracted forms are used less frequently in the IT subcorpus than in the ET subcorpus.

4.2.7 Neologisms and loanwords (N2 operationalisation)

Neologisms and loanwords have been used in a number of other studies as an indicator of normalisation, on the assumption that both these categories of lexical items are an indicator of lexical creativity. If translation does have a linguistically normalising effect, the language of translations will be less creative and more conventional than that of non-translations, which will be reflected in a lower frequency of non-standard lexical items like neologisms and loanwords in comparison to the frequency of such items in non-translations. For the N2 operationalisation, it was hypothesised that neologisms and loanwords would be most prevalent in the NT subcorpus, less in the ET subcorpus and the least in the IT subcorpus.

As with a number of the other operationalisations, the values for this variable were skewed to the right (see Figure 4.23) largely as a consequence of the high number of 0-values.

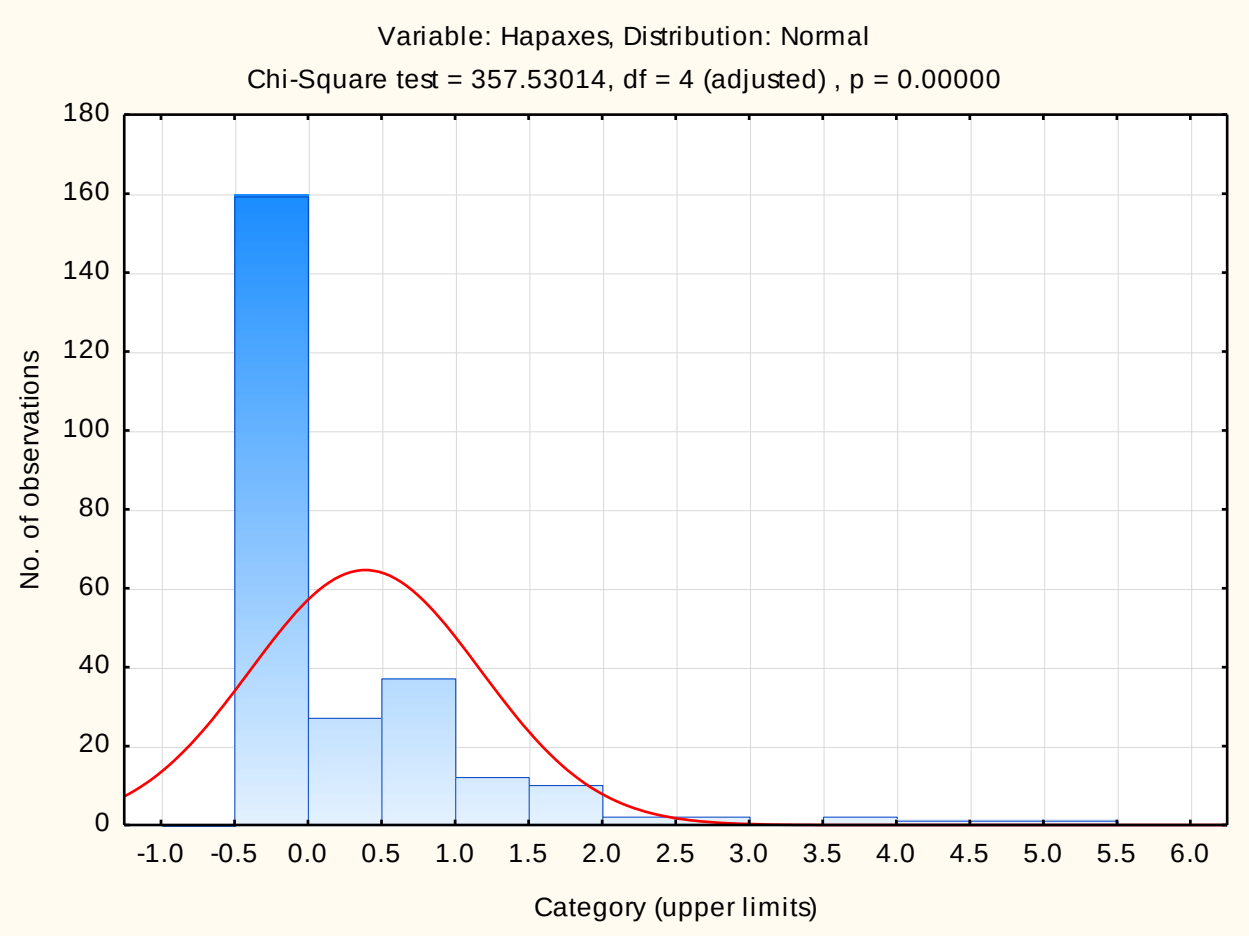

Figure 4.23: Non-normal distribution of values for the variable neologisms and loanwords

The descriptive statistics represented by the boxplot in Figure 4.24 indicates that the median values of the three subcorpora are the same, at 0.0 per 1000 words, which means that in half of the texts in all three subcorpora, neologisms and loanwords do not occur at all. Three-quarters of the texts in the IT subcorpus have a frequency of up to 0.50 creative lexical forms per 1000 words, while in the ET subcorpus three-quarters of the texts have a frequency of up to 0.19 such forms per 1000 words. In other words, there is a greater degree of variability in the IT subcorpus than in the ET subcorpus. In terms of dispersion, the NT subcorpus is more similar to the IT subcorpus (though with an even higher degree of variability), with up to 0.64 creative forms per 1000 words occurring in three-quarters of the texts.

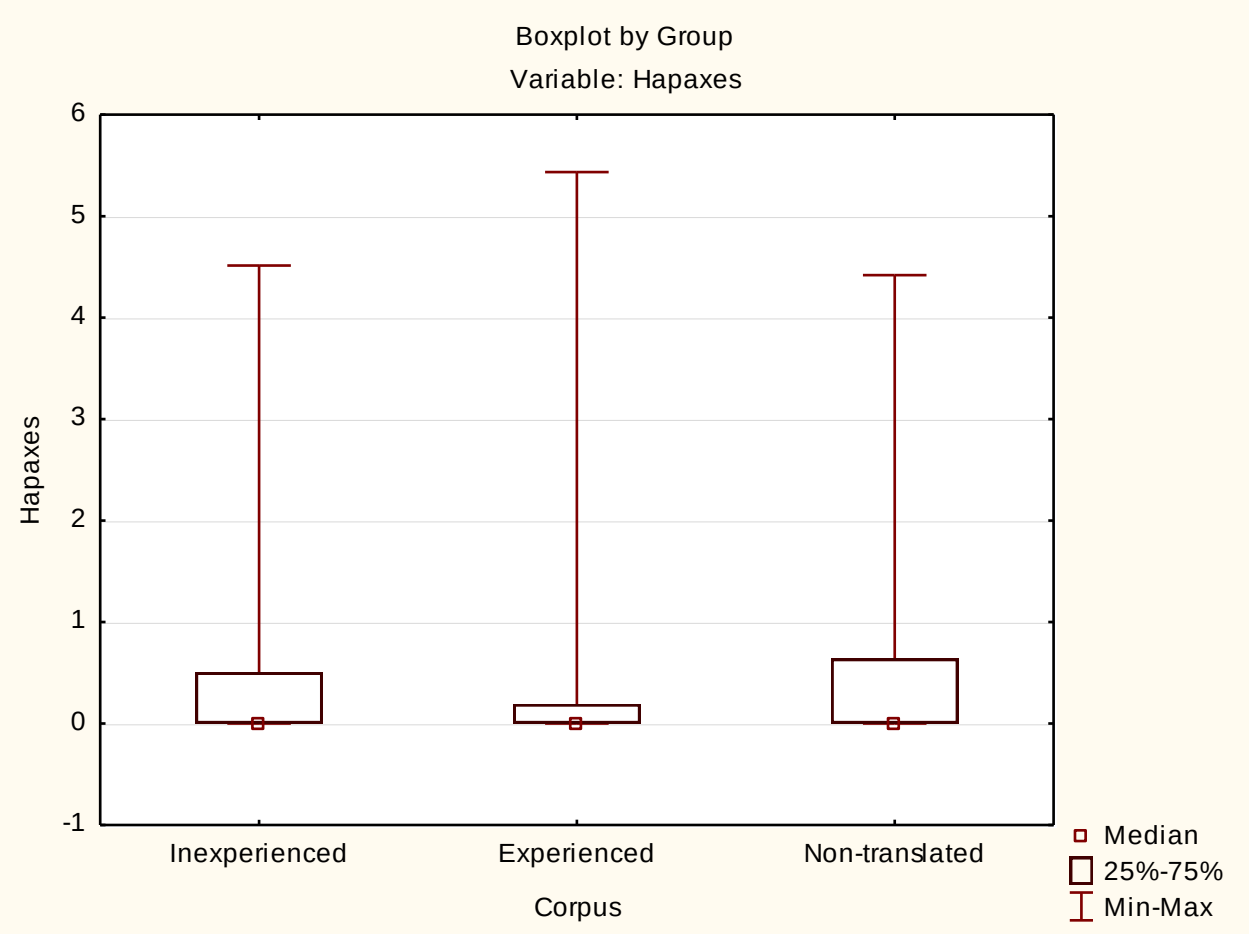

Figure 4.24: Boxplot for neologisms and loanwords

In the IT subcorpus some of the neologisms include words like *downtroddenly*, *doomsaying* and *microbackpack* with loanwords such as *jieyin* (Japanese), *kawaii* (Japanese) and *torregas* (Spanish), whereas the ET subcorpus includes neologisms such as *ashimmer*, *danished* and *nonstate*, with *liewe* (Afrikaans), *bakkie* (Afrikaans) and *kammersängerin* (German) as loanwords, amongst others. A number of the loanwords in the ET subcorpus were from the Germanic languages, especially Afrikaans, whereas the loanwords in the IT subcorpus were from a variety of different languages. In the NT subcorpus, neologisms such as *eyelashed*, *herbivory* and *iconographically* occurred, with loanwords such as *savonniere* (French), *orderél* (Spanish) and *askari* (Arabic).

A Kruskal-Wallis ANOVA shows that the differences between the three subcorpora are statistically significant ($H(2, N=254)=8.05$, $p<0.05$). The NT subcorpus has the highest mean rank (142.06), and therefore the highest median frequency of single-occurrence creative forms. This is followed by the

IT subcorpus, with a mean rank of 126.02, and the ET subcorpus, with a mean rank of 114.40. However, despite the fact that the results are statistically significant, they only partially support the hypothesis formulated for the N2 operationalisation. Even though the translated subcorpora contained fewer instances of neologisms and loanwords than the NT subcorpus, as expected, it was found that the IT subcorpus was lexically more creative than the ET subcorpus.

4.2.8 Register variation (L1 operationalisation)

Register variation was used as an indicator of levelling-out for this study based on the premise that if translated language does demonstrate less variation and therefore greater homogeneity than non-translated language, the operationalisations used for the features of translated language in this study would demonstrate less variability according to register, or less strong register effects, than the non-translated subcorpus. In addition to this, it was predicted that the IT subcorpus would show the least register variation for the features investigated while the ET subcorpus would demonstrate more register-related variation.

The method of data analysis used in this section has been explained in more detail in Section 3.7. A factorial ANOVA is used in each instance to determine whether there is an interaction between the independent variables of corpus and register, which would indicate whether the three subcorpora demonstrate different register-related preferences for the particular linguistic feature under investigation. In instances where such an interaction effect does exist, further investigation is conducted to determine whether the translated subcorpora demonstrate less register variation than the non-translated subcorpus, and to determine whether the IT subcorpus shows the least register sensitivity, as predicted.

4.2.8.1 *That*-omission (E1 operationalisation)

For the *that*-omission ratio the data were found to be highly skewed to the right, mainly because of the high frequency of zero omissions (see Figure 4.1). This precluded the use of a power transformation to normalise the data. As explained in Section 3.7, a factorial ANOVA was performed to determine whether there is a significant interaction for the independent variables corpus and register. However, as is clear from Figure 4.25, there is no statistically significant interaction effect between corpus and register for *that*-omission ratio ($F(8, 180)=1.26$, $p=0.27$), which means that the null hypothesis of no difference cannot be rejected with confidence. In general, the complementiser

that is almost never omitted in the instructional register in any of the three subcorpora and only slightly more frequently in the academic register. Overall, the occurrence of *that*-omission is most varied in the registers of the NT subcorpus while the pattern of *that*-omission appears to be similar for the two translated subcorpora even though the ET subcorpus tends to omit it more frequently, on the whole, than the IT subcorpus. The two translated subcorpora do appear to demonstrate slightly less register variation than the NT subcorpus, but the difference is not significant. Also, the register distribution of this feature follows much the same pattern in the three subcorpora. There is therefore some indication of a levelling-out effect in translation which is more noticeable in the IT subcorpus, but statistical support for this effect is lacking.

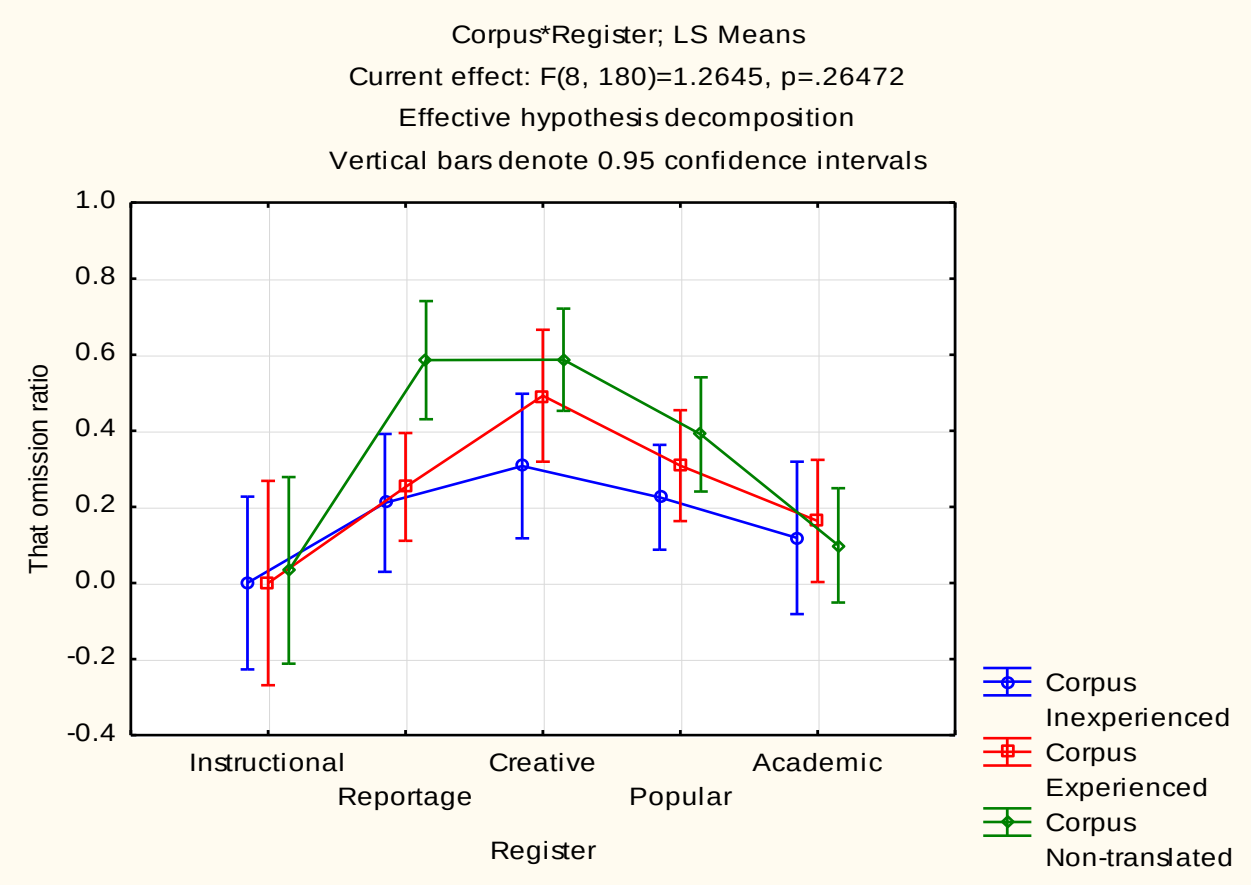

Figure 4.25: ANOVA for *that*-omission ratio in the three corpora and five registers

4.2.8.2 Conjunctive markers (E2 operationalisation)

For the study of conjunctive markers, the data were normally distributed (see Figure 4.9), which means that a factorial ANOVA could be used to investigate the frequency of conjunctives in the three

subcorpora and five registers. The results of the factorial ANOVA are statistically significant ($F(8, 239)=3.11$, $p<0.05$), as is represented in Figure 4.26. In the creative writing and popular writing registers, the IT subcorpus has a surprisingly higher incidence of conjunctive markers as opposed to the instructional and academic registers where it has the lowest occurrences. In the reportage register it has a similar occurrence to the ET and NT subcorpora. In general, it also has the most varied pattern of the three subcorpora. It is interesting to note in this regard that the higher incidence of conjunctive markers is noticeable only in the registers that are intrinsically more informal in nature than those which are more formal, which suggests that inexperienced translators only explicitate by using conjunctive markers in the more informal registers and are therefore less sensitive to register preferences for these informal registers. It may well be that inexperienced translators consciously or unconsciously overcompensate in the translation process due to their inexperience.

In general, the ET and NT subcorpora used conjunctive markers in similar ways across the five registers, even though the ET subcorpus used them slightly more in the reportage, creative writing and academic writing registers. A similar pattern is therefore evident for these two subcorpora in terms of the distribution of conjunctives across the five registers. This may be taken to indicate that the register sensitivity in the ET subcorpus is similar to the NT subcorpus in terms of this feature, even though the overall frequency is slightly higher in the ET subcorpus in some registers.

In sum, the varied pattern of the IT subcorpus and the similarity between the ET and NT subcorpora can be taken to indicate that translation-related levelling-out in the use of conjunctive markers across the five registers in the translated subcorpora does not occur. It does appear, however, that the inexperience of translators plays a role, with the IT subcorpus demonstrating a noticeably different pattern for the distribution of this feature compared to the other two subcorpora. The suggestion is that inexperienced translators misjudge the appropriateness of the addition of conjunctive markers particularly in more informal registers.

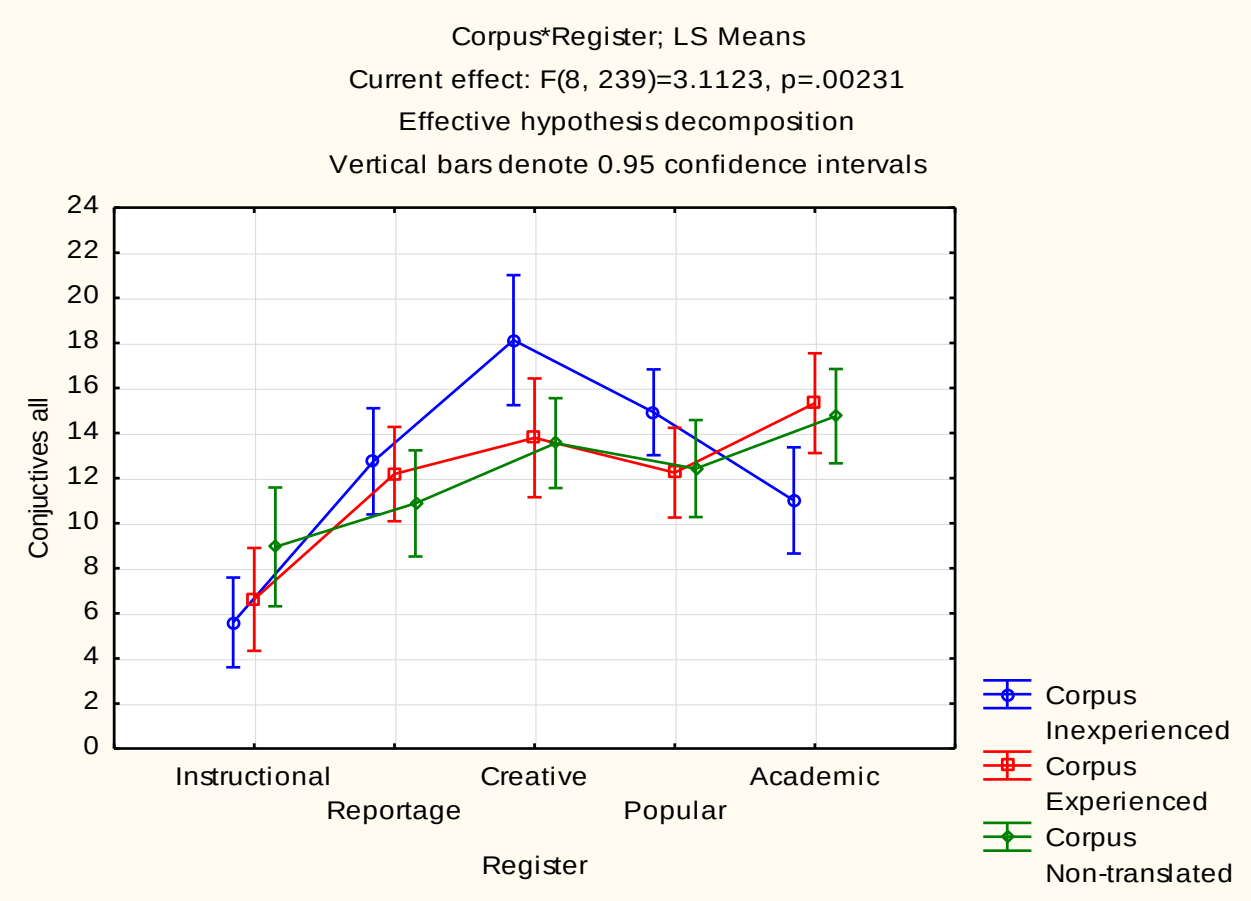

Figure 4.26: ANOVA for all conjunctives in the three subcorpora and five registers

4.2.8.3 Type-token ratio (S1 operationalisation)

For standardised TTR, a Box-Cox transformation could be performed to normalise the non-normally distributed data (see Figure 4.11). Figure 4.27 provides information regarding the transformation of the data.

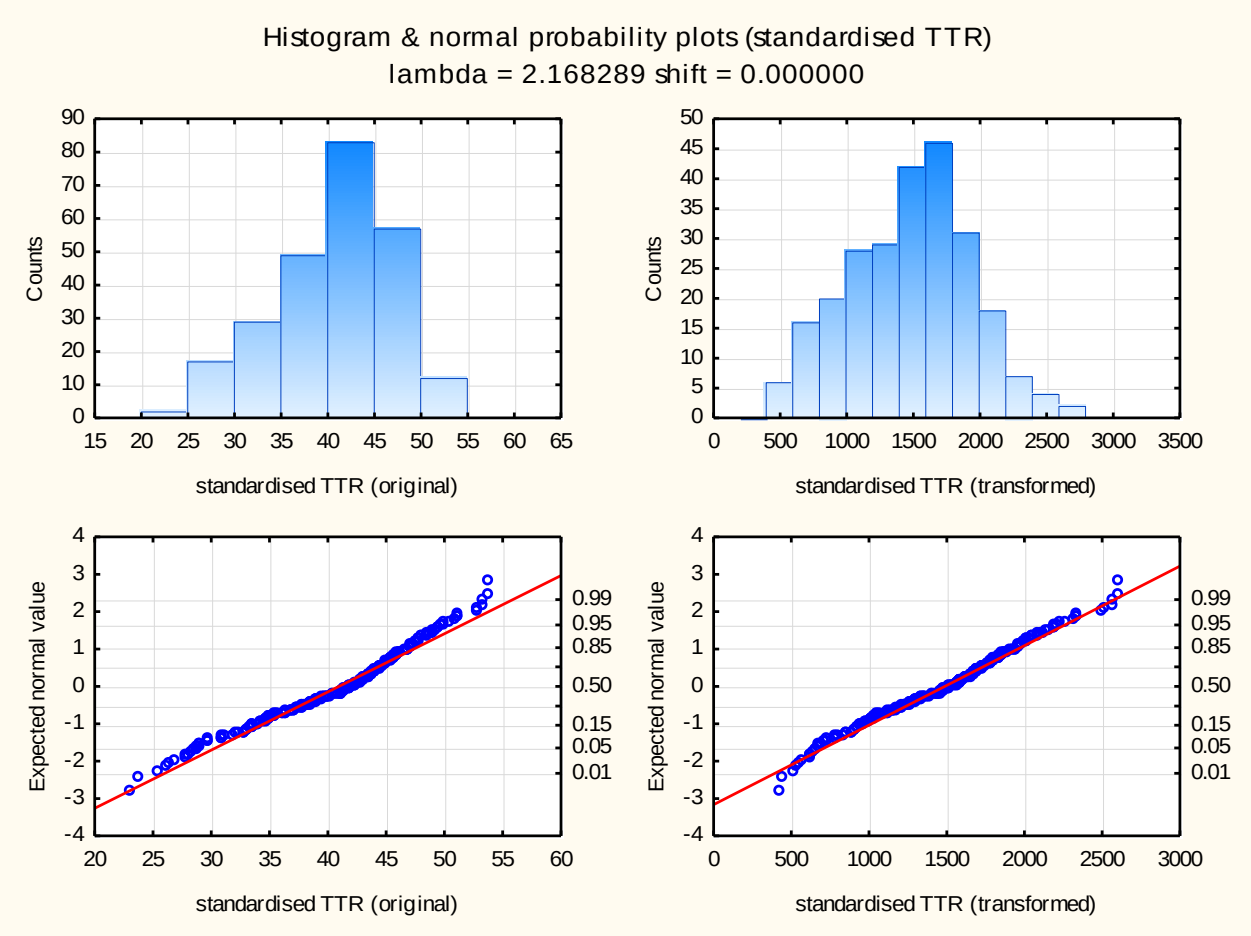

Figure 4.27: Box-Cox transformation of data for standardised TTR

A factorial ANOVA was performed using the transformed data. The results demonstrate a statistically significant interaction effect ($F(8, 234)=3.57$, $p<0.001$) for the variables corpus and register, as indicated by Figure 4.28. On the whole, the pattern of the standardised TTR values across the different registers appears to be quite similar between the NT and the ET subcorpora, even though standardised TTR values for registers in the ET subcorpus are generally lower overall than the standardised TTR values for registers in the NT subcorpus. For the IT subcorpus, the pattern is somewhat different. With the exception of the instructional register, the IT subcorpus has lower standardised TTR values than the other two subcorpora throughout. In the creative writing register, where the other two subcorpora tend to have a fairly high standardised TTR (compared to the values in the other registers), the IT subcorpus has a lower standardised TTR (compared to the values in the other registers for this subcorpus). It therefore appears that lack of experience means that inexperienced translators struggle to attain the lexical diversity that is typical of creative writing. In

the popular writing register, however, where the NT and ET corpora have similar values for standardised TTR in the creative and popular writing registers, the IT subcorpus has a noticeably higher standardised TTR for the popular register as compared to the creative writing register. Once again, it appears that inexperienced translators have more difficulty with informal registers. However, despite the statistically significant interaction effect for corpus and register, there is no real evidence for a levelling-out effect, with all three subcorpora demonstrating some register variation.

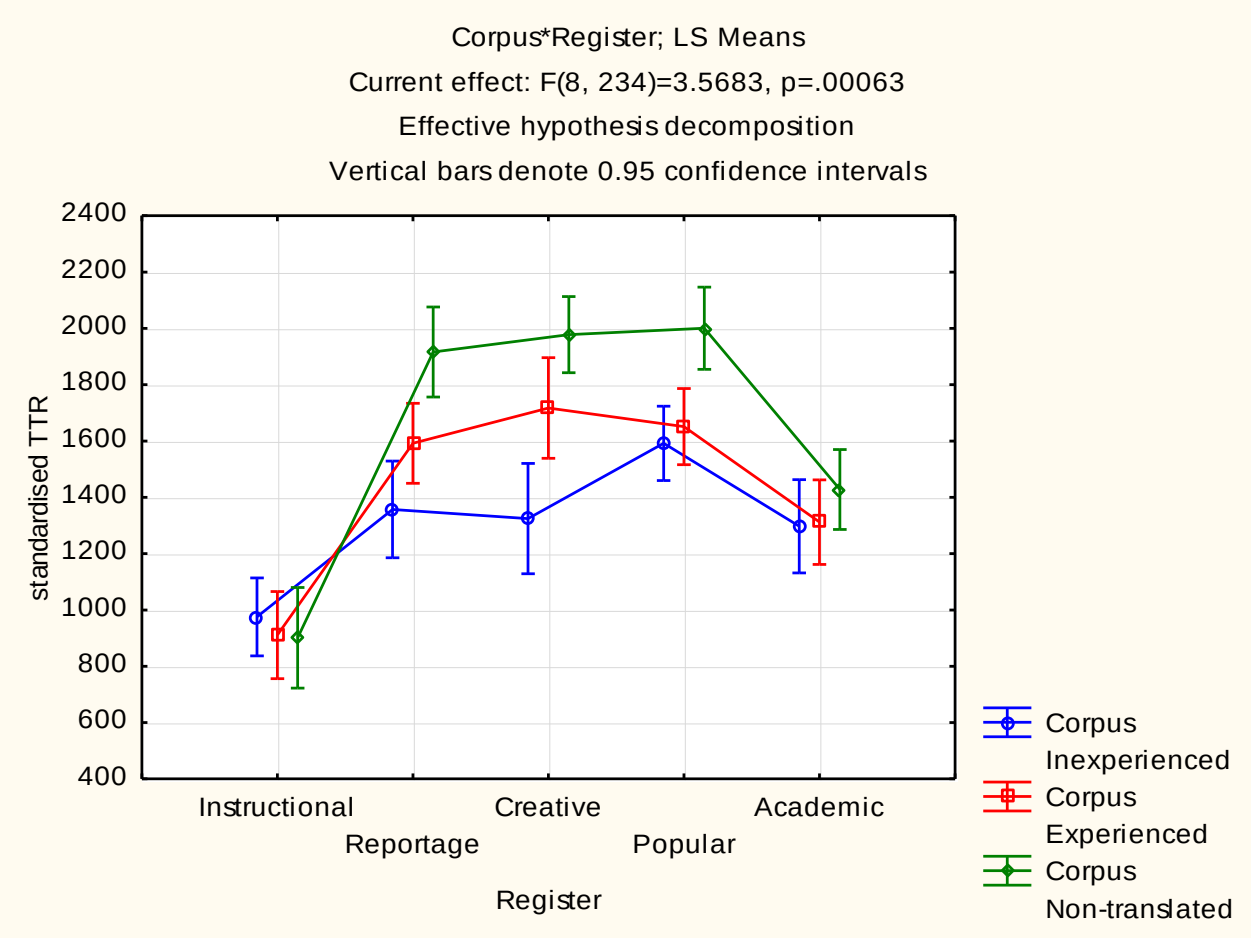

Figure 4.28: ANOVA for standardised TTR in the three subcorpora and five registers, using transformed data

4.2.8.4 Readability score (S2 operationalisation)

For the readability score, the data were normally distributed (see Figure 4.13) and a factorial ANOVA was performed. Figure 4.29 shows that there is an obvious register effect, which is independent of translated status or degree of experience. There is no significant effect for the interaction between the variables of corpus and register ($F(8, 239)=0.60$, $p=0.78$) and the null hypothesis (H_0) of no difference cannot be rejected with confidence. All three of the subcorpora display a similar pattern for the five registers, and their values are relatively similar to one another. The highest scores can be noted in the creative writing and popular registers, which are the most informal of the registers in the corpus, and the registers which would be expected to have the highest degree of readability. The instructional and academic registers have the lowest scores, which means that they are the most difficult to read, again in line with what one might expect for formal registers like these. Because of this strong register effect, there is no indication of levelling-out occurring in either of the translated subcorpora.

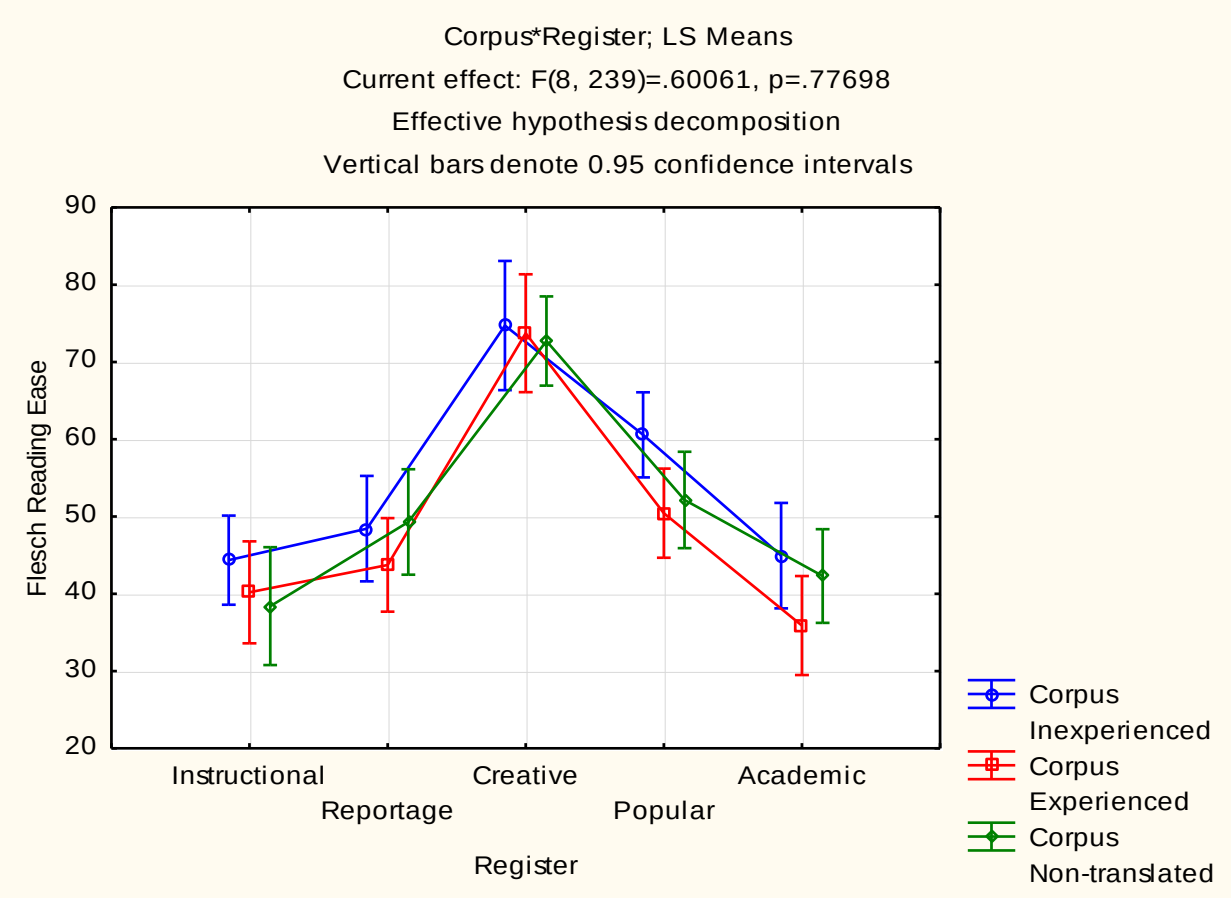

Figure 4.29: ANOVA for Flesch Reading Ease score in the three subcorpora and five registers

4.2.8.5 Word length (S3 operationalisation)

As the data for this feature were normally distributed (see Figure 4.15), a factorial ANOVA could be performed to test whether word length demonstrates an interaction effect for corpus and register. The results just fail to reach statistical significance ($F(8, 239)=1.84$, $p=0.07$), which means that there is some evidence for differences in register distribution of this feature across the three subcorpora. From Figure 4.30 it can be seen that the register pattern for mean word length is very similar for the three subcorpora. The creative writing register has the shortest mean word length in all three of the subcorpora, while the more formal instructional and academic registers have the longest word lengths in all three subcorpora. The NT subcorpus, however, behaves somewhat differently in the instructional and reportage registers, compared to the two translated subcorpora, in that the values for mean word length are very similar for these two registers in the NT subcorpus, while the two translated subcorpora both have a higher mean word length in the instructional than the reportage register. However, there is no evidence that translation causes a levelling-out of word length across registers.

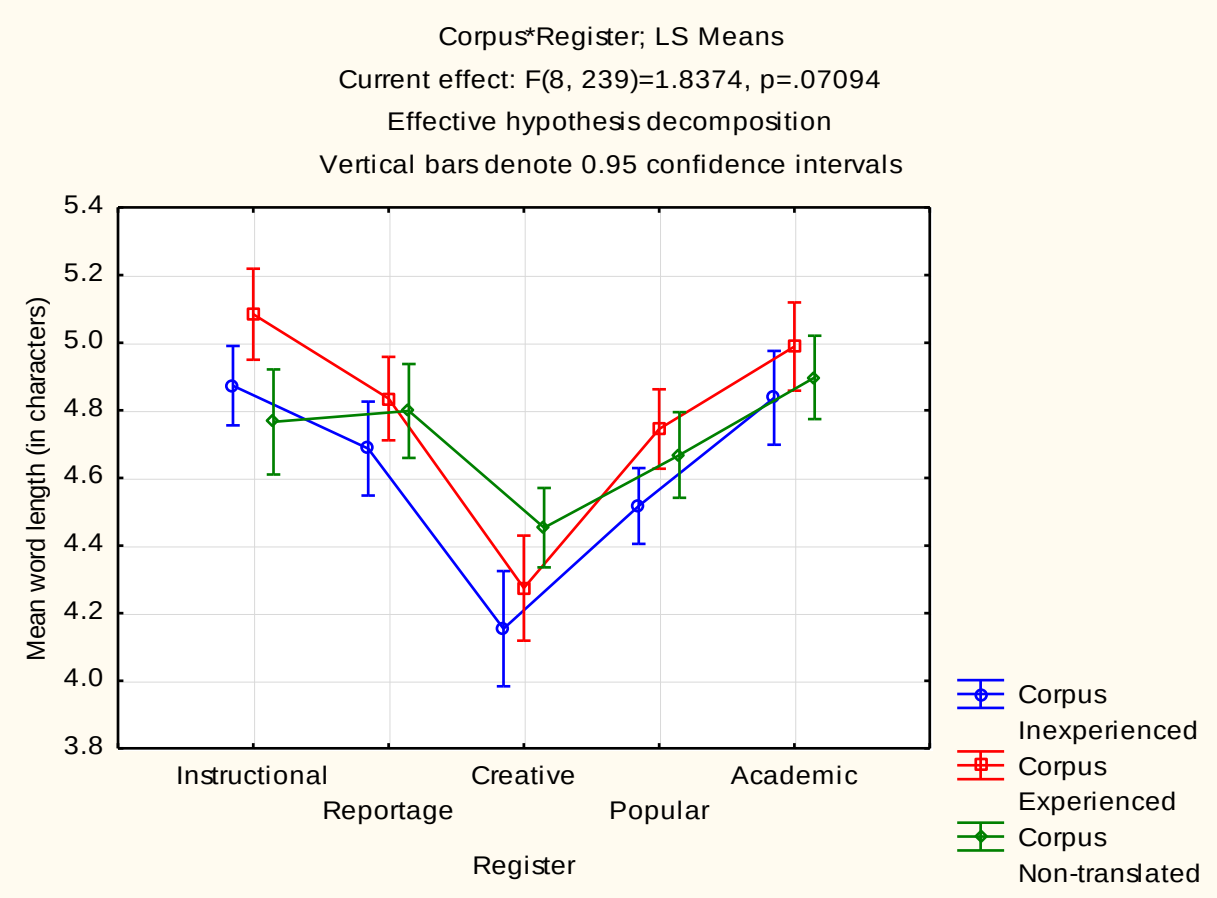

Figure 4.30: ANOVA for word length in the three subcorpora and five registers

4.2.8.6 Contraction ratio (N1 operationalisation)

Data for the variable contraction ratio (including both *not*-negation and verb contractions) were not normally distributed (see Figure 4.21). As a consequence of the high number of 0-values for this variable, the data could also not be transformed. Since there were sufficient data points, a factorial ANOVA was performed regardless of the non-normality of the data. The results show no significant interaction effect for corpus and register ($F(8, 237)=1.60$, $p=0.13$) which means that the null hypothesis of no difference cannot be rejected with confidence.

Figure 4.31 shows that the three subcorpora demonstrate very similar register preferences for the use of contracted forms. The IT and ET subcorpora demonstrate values in much the same range, which peak in the creative writing register, as expected. The NT subcorpus has similar values for the instructional, reportage and academic registers when compared to the translation subcorpora, but has a noticeably higher value for the creative writing register, and a somewhat higher value for the popular writing register. While there is a suggestion, visually, of some degree of levelling-out of register differences in the translated subcorpora, the statistical analysis does not offer enough support to meet the criterion level of a chance result.

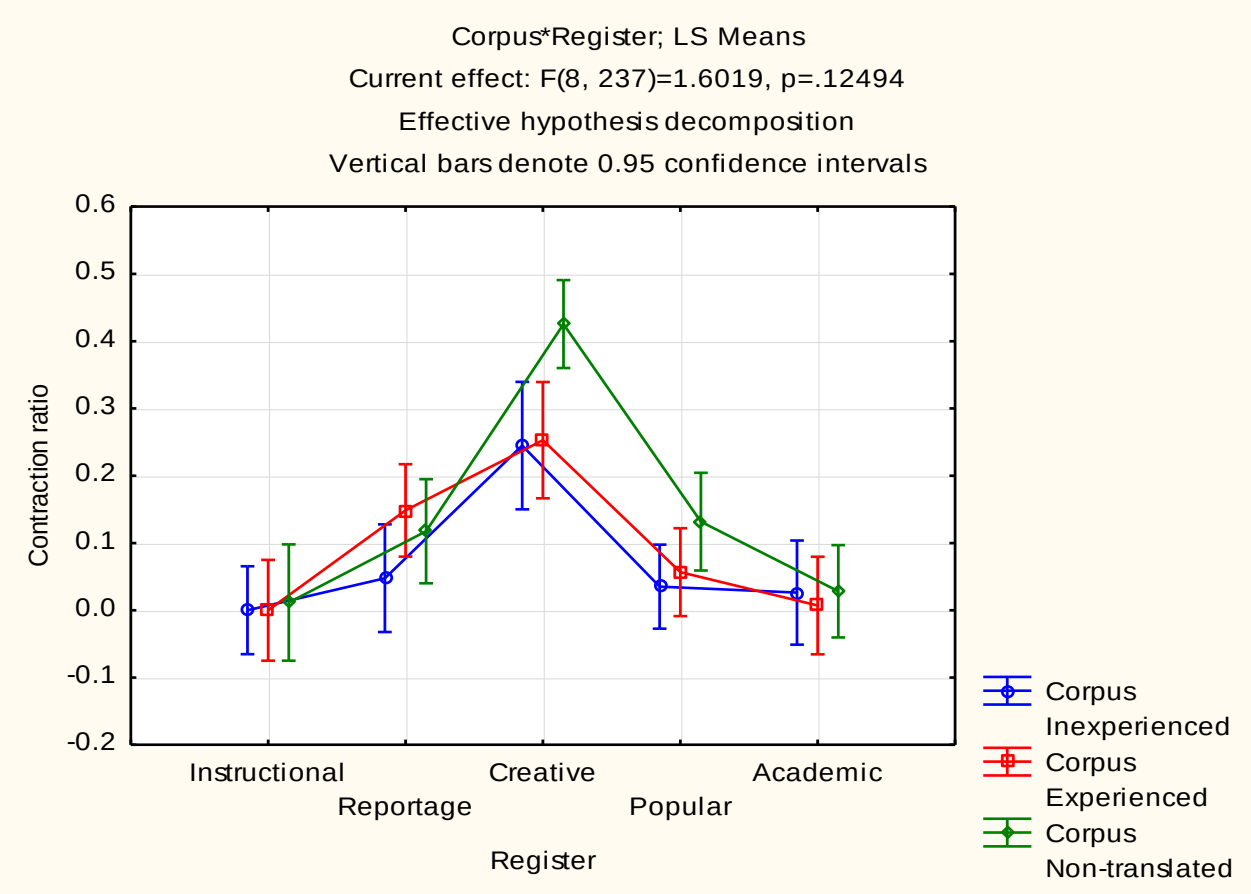

Figure 4.31: ANOVA for contraction ratio in the three subcorpora and five registers

4.2.8.7 Neologisms and loanwords (N2 operationalisation)

As is the case for the N1 operationalisation, the data for the N2 operationalisation were not normally distributed (see Figure 4.23) and could not be normalised by a power transformation, as a consequence of the high number of 0-values. Regardless of the non-normality of the data, a factorial ANOVA was performed, since there were sufficient data points for the analysis.

The results of the factorial ANOVA show that there is no interaction for register and corpus ($F(8, 239)=1.40$, $p=0.20$) and, as such, the null hypothesis of no difference cannot be rejected with confidence. From Figure 4.32 it can be seen that there clearly is no interaction and no effect at all for this feature in the three subcorpora and across the five registers, due to random fluctuations in the data. There is therefore also no support for the levelling-out hypothesis.

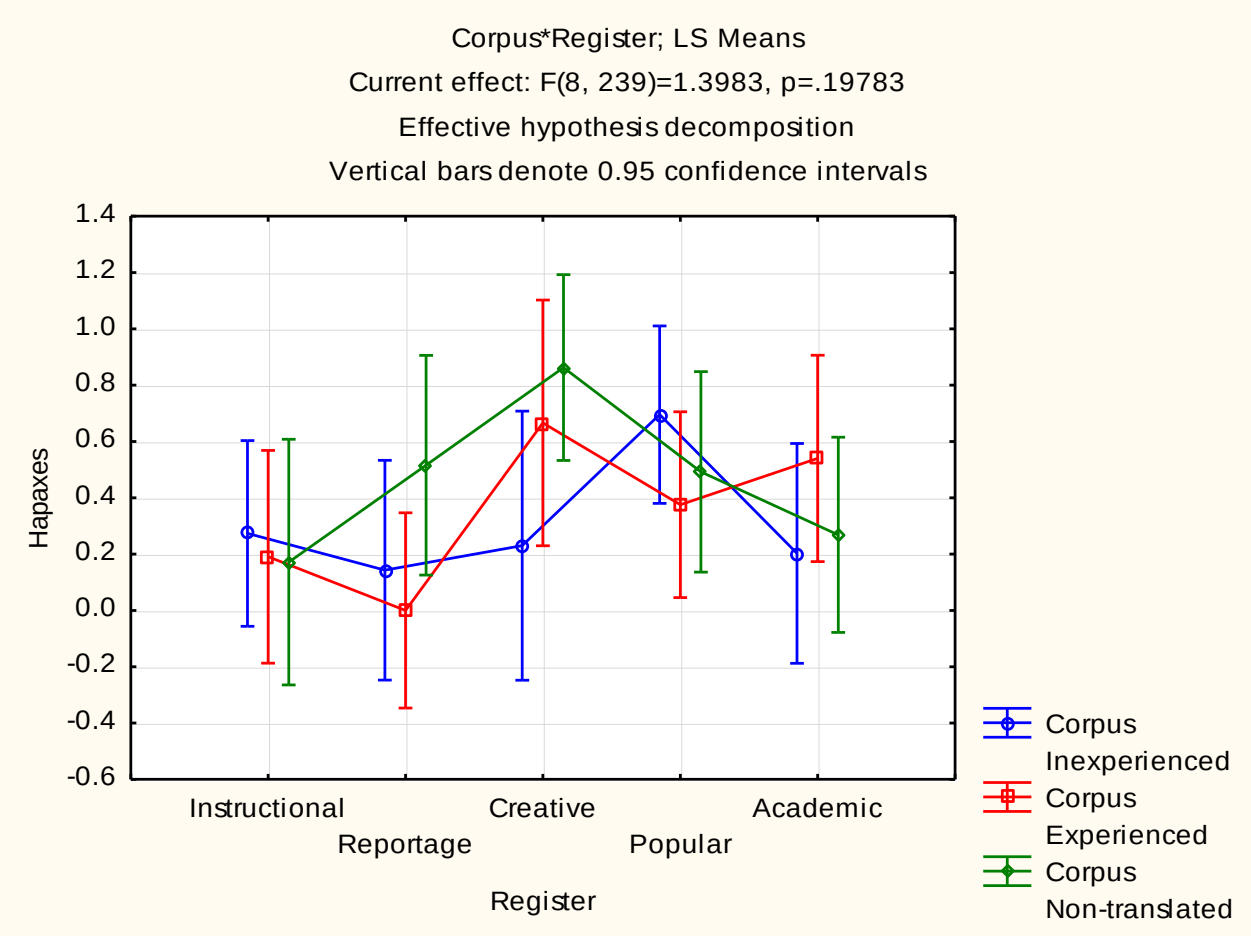

Figure 4.32: ANOVA for neologisms and loanwords in the three subcorpora and five registers

4.3 Interpretation of findings

4.3.1 Explicitation

As discussed in Section 2.3.3.1(a), translation-related explicitation has been interpreted and operationalised in a variety of ways. Interpretations of explicitation frequently involve concepts such as redundancy and self-referentiality. Explicitation may be reflected in the inclusion of material that is both formally and semantically redundant, and may also be evident in the increased self-referentiality of a text, as a gloss or contextual information is added.

Two operationalisations were selected for this study to establish not only whether there is support for the claim that translations are more explicit than non-translations, but also whether noteworthy explicitation differences are evident between the target texts produced by experienced and inexperienced translators. The first of the two operationalisations focused on the omission of the

complementiser *that* to determine whether the translation subcorpora are more explicit, in that they tend to make grammatical relations clearer by using the optional *that* more with reporting verbs, than the NT subcorpus. In keeping with the explicitation hypothesis, it was expected that the translated subcorpora would have a lower omission ratio of the complementiser *that* than the NT subcorpus, but also that the translators of the ET subcorpus would have been motivated to explicitate more frequently than the translators of the IT subcorpus, so that a lower omission ratio was expected in the ET than in the IT subcorpus. This hypothesis for the E1 operationalisation was based on the assumption that experienced translators would be more inclined to avoid the risk of failed communication.

While the first operationalisation focuses on formal completeness, the second of the two explicitation operationalisations is based on the argument that explicitation may involve making the semantic relations between propositions clearer in order to help the reader with the interpretation of the text or to minimise ambiguity. Such instances of explicitation may be flagged by the use of conjunctive markers, since these markers can help to make the link between ideas and propositions clearer and more explicit. In line with the explicitation hypothesis, it was hypothesised that the translated subcorpora would demonstrate higher frequencies of conjunctive markers than the NT subcorpus. In addition it was foreseen that the ET subcorpus would show a higher frequency of the use of conjunctive markers than the IT subcorpus, as experienced translators would be more aware of fewer shared cultural references between the source and target audiences and would therefore be more likely to restate or introduce information to help the target audience with their interpretation of the text.

The findings indicate statistically significant differences between the three subcorpora for only the first of the two explicitation operationalisations (E1); however, the results were not entirely as expected. The NT subcorpus omitted that the most frequently of the three subcorpora, as hypothesised, and there is therefore convincing evidence for a translation-specific explicitation effect. As far as the effect of expertise is concerned, it was found that inexperienced translators practically never omit the *that* complementiser, which essentially means that inexperienced translators are most prone to explicitation, contrary to the hypothesis that experienced translators would explicitate more than inexperienced translators.

The possible confounding effect of source language on the findings needs to be considered. Languages like French, Italian and Czech require a subordinator following reporting verbs (Becher, 2010b:5). Thus, if texts translated from languages such as French were overrepresented or dominant in the IT subcorpus compared to the ET subcorpus, it may have influenced the results, leading translators to include the optional complementiser because of the priming effect of the subordinator that is presented in the source text. However, in the corpus used for this study, texts translated from French represent 7.86% of the IT subcorpus and 11.28% of the ET subcorpus, whereas translations from Italian represent 1.07% of the IT subcorpus and 1.03% of the ET subcorpus (see Section 3.3.2). No translations from Czech were used in the custom-built corpus. It therefore appears unlikely that source-language influence could account for the differences between the IT and ET subcorpora.

The greater degree of formal explicitness in the IT subcorpus as compared to the ET subcorpus therefore requires an alternative explanation. Even though it has been hypothesised that experienced translators will be more risk-averse than inexperienced translators, it may be that student translators are risk-averse not because they fear that their audience will not understand the translation, but because they are attempting to meet what they perceive as the standard for grammatically correct writing expected by them from their teachers. However, it should be kept in mind that a large proportion of the IT subcorpus is, in fact, made up not of student translations, but translations done for pleasure by laypersons, and this explanation therefore appears unlikely. A more likely explanation is that inexperienced translators are more inclined to always opt for the full form because they are so inexperienced at translation that they tend to explicitate the meaning of the source text for themselves more consistently, which is then reflected in their translations.

Another plausible reason is that inexperienced translators are not as sensitive to register preferences for *that*-omission as experienced translators are and therefore cannot effectively judge in which registers it is acceptable to omit the complementiser *that*. The analysis of register variation for this feature presented in Section 4.2.8.1 shows that the NT subcorpus had the most varied pattern for *that*-omission of the three subcorpora, and particularly omitted the complementiser the most in the reportage and creative writing registers. The IT subcorpus showed less variation in the distribution of the feature of *that*-omission across the different registers, indicating inexperienced translators' lack of register sensitivity (see Figure 4.25). The ET subcorpus demonstrated slightly more register variation for this feature than the IT subcorpus, suggesting greater register sensitivity on the part of

experienced translators. The register sensitivity associated with translation expertise, therefore, provides a possible explanation for the findings.

The findings for the E2 operationalisation showed no statistically significant difference between the three subcorpora. For all three categories of conjunctive markers selected for the study, namely markers of elaboration, extension and enhancement, the frequency with which these markers are used in the three subcorpora are very similar. Explicitation at the semantic level is therefore not a distinctive feature of translated language, nor is there any difference in how translators with different levels of expertise make use of conjunctive markers to mark and explicitate the propositional relationships in texts.

The analysis of the register distribution of conjunctives across the three subcorpora does, however, suggest that inexperienced translators have some difficulty in judging register-related preferences for the use of conjunctives. As discussed in Section 4.2.8.2, the ET and NT subcorpora tend to use the conjunctive markers in a similar way across the various registers. Experienced translators, in other words, are sensitive to the register-related conventions for use of these conjunctive markers in non-translated texts and adjust their translations accordingly. The IT subcorpus, on the other hand, demonstrates a somewhat different pattern in its use of conjunctive markers, with a noticeably higher occurrence of these markers in the informal registers of creative writing and popular writing, suggesting the difficulty inexperienced translators have in gauging the register-appropriate use of conjunctives.

In sum, for the feature of explicitation it appears that (a) translators do tend to explicitate for formal completeness, but not for semantic addition and (b) in the case of formal completeness, inexperienced translators are more consistently prone to explicitation than experienced translators. It appears likely that this greater tendency towards explicitation in the work of inexperienced translators is the consequence of a cognitive explicitation process which is exacerbated for inexperienced translators because of their lack of translation practice, or of an underdeveloped sense of register-related preferences for *that*-omission, or, most probably, a combination of the two factors.

4.3.2 Simplification

According to theories of simplification in translation, translators simplify their translations either consciously, to help the target audience with their interpretation of the text, or unconsciously, as a consequence of the cognitive processes involved in translation, which constrain the range of linguistic options that translators make use of, primarily because of the influence of the source text. Three operationalisations were chosen to test whether translated texts are, indeed, simpler than originally produced texts, and also to establish if there is an observable difference between the translations of experienced and inexperienced translators as a consequence of the use of different translation strategies in the two groups.

One of the predictions of the simplification hypothesis is that translations will display a more limited vocabulary range than non-translations. Standardised TTR, a measure of lexical diversity, is used as the first operationalisation for simplification. For this operationalisation, it was hypothesised that the translated subcorpora would demonstrate a lower standardised TTR than the non-translated subcorpus, and that the IT subcorpus would have a lower standardised TTR than the ET subcorpus. It was expected that the IT subcorpus would display a lower TTR as inexperienced translators still need to improve or develop their language skills and knowledge, which implies that they would have a more restricted vocabulary than both more experienced translators and original text producers.

For the S2 operationalisation, the Flesch Reading Ease score was chosen as a measure of simplification, as this score serves to measure the level of skill needed by readers to make sense of a text. It was hypothesised that the translated subcorpora would score higher on the Flesch Reading Ease Test than the NT subcorpus (where a higher score indicates greater reading ease). Additionally, it was thought that the ET subcorpus would show a higher degree of readability than the IT subcorpus, as the experienced translators would most likely be more sensitive to audience needs and aim to make the text easier for the reader to understand.

Readability indices use a combination of sentence length and word length in calculating the readability score of a text. The S3 operationalisation, word length, may therefore be seen as an additional validation of the findings for the S2 operationalisation, without the effect of sentence length (which may also be tied to explicitation). It was expected for the S3 operationalisation that the

ET subcorpus would have a lower average word length than the NT subcorpus as, according to the simplification hypothesis, experienced translators will try to simplify the message of their translations for their target audience. In addition, simplification is also believed to occur as a consequence of cognitive processes involved in the translation process and as such would also occur in the IT subcorpus (but to a lesser degree than in the ET subcorpus), which would be reflected in a mean word length for the IT subcorpus somewhere between the ET and NT subcorpora.

Findings for the S1 and S3 operationalisations demonstrate a statistically significant difference between the three subcorpora, while the results of the S2 operationalisation show a difference close to statistical significance. The findings confirm the hypothesis formulated for the S1 operationalisation, as the IT subcorpus does have the lowest TTR, making it the subcorpus with the smallest degree of lexical diversity; the NT subcorpus the highest, with the greatest range of vocabulary; and the ET subcorpus somewhere in between. The findings for the S2 and S3 operationalisation were contrary to what was expected. As predicted, the IT subcorpus is the easiest to read (and therefore the least complex), with the highest average readability score and the shortest average word length. However, the ET subcorpus has the lowest average readability score and the longest mean word length, which means that this subcorpus has the greatest degree of complexity of the three subcorpora, as measured by these operationalisations. The complexity of the NT subcorpus is somewhere in between the IT and ET subcorpora in terms of both the S2 and S3 operationalisation.

The findings for the S1 operationalisation provide support for the original hypothesis. The lexical diversity of the NT subcorpus is higher than that of both translated subcorpora, suggesting a reduction of lexical range, and therefore simplification, as a consequence of translation. It was anticipated that the ET subcorpus would have a higher standardised TTR than the IT subcorpus, because experienced translators are expected to have developed a language competence which involves syntactic, morphological and vocabulary knowledge (see Section 2.2.2.1). This broader vocabulary range is related to the quality of translation as, according to Scarpa (2006:166), better translations will "have a greater lexical variation, i.e. display a higher use of synonyms and a lower use of repetition". The low standardised TTR for the IT subcorpus is accounted for by inexperienced translators' still developing language competence. In addition, experienced translators are likely to be more aware of the constraining effects of the source text on lexical diversity in the target text, and develop compensatory strategies (e.g. the conscious use of synonymy) – which inexperienced translators have

not yet had the opportunity to develop due to their lack of professional experience. This would, in other words, form part of the linguistic or language component of expertise discussed in Section 2.2.2.1. Furthermore, the greater lexical range in the work of experienced translators as opposed to inexperienced translators may also be accounted for from a cognitive-linguistic perspective. As discussed in Section 2.2.2.3, experienced translators are able to work on the source text's communicative cues in a more effective manner, as they can distinguish between the contextualised meaning of words and the dictionary meaning while bridging the gap between conceptually, procedurally and contextually encoded information as well.

For the other two simplification operationalisations, findings did demonstrate that the IT subcorpus is the most simplified, morphologically and syntactically, which is most likely related to the less-developed language skills of inexperienced translators. Contrary to expectations, however, the ET subcorpus was the most complex along the parameters of sentence length, word length, and overall readability, when compared to the IT and NT subcorpora. This, therefore, does not provide support for the hypothesis that simplification is an inherent trait of translated texts, begging possible explanation for the findings. According to theorists such as Gale and Church (1993:75) and Rosner (2004:83) longer sentences in the source text are usually translated into longer sentences in the target text, whereas shorter sentences are usually translated as shorter sentences. This therefore raises the possibility that the source texts of the translations in the ET subcorpus may have had longer sentence lengths to begin with, which may have primed the longer sentence lengths in the target texts. In this respect it should also be kept in mind that the majority of texts in the IT subcorpus were produced by laypersons who mainly translated for pleasure or for practice. This may have had an influence on their text selection, leading them to select easier texts with shorter sentence lengths. Similarly, it may be that the student translations in the IT subcorpus simply had "easier" source texts than the translations in the ET subcorpus, which contains authentic commissioned translated texts.

Another potential explanation is the possibility that the experienced translators increased sentence and word length in conforming to perceived norms for written text production, which in essence means that normalisation may have played a role. According to Greenfield and Subrahmanyam (2003:722) written language often uses longer sentences overall with a more formal and conventional grammar and vocabulary. It is therefore a possibility that the experienced translators normalised to a written standard and consequently exaggerated and overrepresented the features associated with written

English, resulting in longer sentence lengths and longer words. Original text producers are not expected to be as limited by these written conventions, and inexperienced translators are assumed to be less aware of standards for written language.

Something else to keep in mind is the tension in the relationship between explicitation and simplification as the two concepts are believed to run counter to one another, as discussed in Section 2.3.3.2.(a). This is because explicitation is expected to lead to longer sentences, as a consequence of addition, while simplification would shorten sentences. Thus, if longer sentences are a feature of explicitation it may well mean that the occurrence of the longer sentences in the ET subcorpus is caused by explicitation. However, for this particular study, the operationalisations chosen do not provide support for the notion that explicitation accounted for the longer sentences. This is because the IT subcorpus used the complementiser *that* more frequently than the ET subcorpus (which, in essence would mean that the sentences of the ET subcorpus would have been comparatively shorter in this instance), while there is no statistical support for the supposition that experienced translators add conjunctives to help readers with the interpretation of their texts. Therefore, explicitation does not serve as possible explanation in this case.

In sum, the results of the simplification operationalisations do not provide straightforward evidence of simplification. Translation does appear to constrain lexical range, simplifying texts in this respect, and for inexperienced translators this effect is stronger than for experienced translators. However, in terms of morphological and syntactical complexity, it appears that linguistic skill, expertise and experience play a more important role in determining the complexity of a text than translation. The translations of the inexperienced translators evidently simplified the most according to all three operationalisations, but the translations of experienced translators contained the longest sentence and word lengths overall, suggesting that prototypical translation does not necessarily have a simplifying effect on the morphological and syntactic levels.

4.3.3 Normalisation

According to the normalisation hypothesis, translators are often pressured by economic or socio-cultural constraints to conform to the conventions of written language that are applicable to the target language, and even to exaggerate or overrepresent these conventions. Some areas that might be influenced by these constraints and conventions include grammatical structure, punctuation, lexis and

idiomaticity, amongst others, resulting in target texts which are less marked, less creative and more conformist than non-translations produced in the same languages (see Section 2.3.3.3, 3.5.3). In addition to the influence of translation norms, normalisation is also believed to be caused by unconscious cognitive processes that come into play during translation, because of a gravitational pull between what is known as the category prototype and the highest-level schema which results in an overrepresentation of typical features of the target language (see Section 3.5.3). Two operationalisations were chosen to determine not only whether normalisation is more prevalent in translated texts, but also to establish how the translations produced by experienced and inexperienced translators differ from one another. The first of the two operationalisations concentrated on full forms versus contracted forms (N1 operationalisation) and the second focused on neologisms and loanwords (N2 operationalisation).

The use of full rather than contracted forms may be regarded as a measure of the degree of formality of a text, since the use of contractions is discouraged in a variety of written registers as a consequence of its association with spoken and informal registers. According to the normalisation hypothesis, a variety of different pressures involved in the translation process may coerce translators to normalise to a conventional interpretation of the conservative written standard. Based on this idea, it was hypothesised that the translated subcorpora would demonstrate a higher frequency of full forms as opposed to contractions, whereas the subcorpus of non-translated texts would have a higher frequency of contracted forms. However, as it is unlikely that inexperienced translators experience the same pressures as experienced translators, and since their lack of experience would mean that they have had less opportunity to develop sensitivity to register variation, it was hypothesised that their translations would be more stylistically homogenous than those produced by experienced translators. Thus, it was predicted that the IT subcorpus would show the highest degree of full forms as inexperienced translators would lack the experience to know when contractions are acceptable or preferable for a specific register.

As normalisation is believed to result in target texts which are less marked, less creative and more conformist than non-translations produced in those same languages, it was decided to use neologisms and loanwords which occur once in the subcorpora as markers of creativity to determine the degree to which translations are more conservative in their lexical creativity compared to non-translations. It was hypothesised that the NT subcorpus would have the highest incidence of neologisms and

loanwords that occur as hapax legomena (Kenny, 2001:130) when compared to the two translation subcorpora, since original text producers do not experience the pressure to conventionalise to the same degree as translators do. In addition, it was hypothesised that the ET subcorpus would have a higher frequency of these creative words (which also represent a greater willingness to take linguistic risks) than the IT subcorpus as, according to theories of translation expertise, creativity forms part of translation expertise (see Section 3.6.3), and experienced translators would therefore be more adept at applying their creative skills than inexperienced translators.

For both of these operationalisations, the differences between the three subcorpora were found to be statistically significant. It was established that not only do the translated subcorpora use contractions less than the NT subcorpus, but the IT subcorpus shows an even lower frequency of contractions than the ET subcorpus, as expected. This finding may be taken as indicative of the idea that inexperienced translators are, on a whole, less inclined to take risks with non-standard forms, and are also less sensitive to register requirements than experienced translators. According to Wyatt (2007:138), inexperienced translators struggle to make register-sensitive choices as they pay more attention to accurate meaning transfer and, consequently, disregard information about register and style. Even though the ET subcorpus uses contractions more frequently than the IT subcorpus, the frequency of contractions is more similar in the two translated subcorpora than in the NT subcorpus. In the case of the experienced translators, it is likely that the relatively low frequency of contractions is not primarily the consequence of a lack of register sensitivity, but rather of experienced translators' tendency to gravitate more towards a conservative written standard that may be more acceptable to the target audience. These findings therefore provide evidence for the normalisation hypothesis overall.

The frequency of neologisms and loanwords also demonstrates statistically significant differences among the three subcorpora, although though not quite as expected. The translated subcorpora do use single-occurrence creative forms less frequently than the NT subcorpus, providing support for the hypothesis of translation-related normalisation. However, the results indicate that the IT subcorpus has a higher frequency of such forms than the ET subcorpus, contrary to expectation. This finding may be explained by prototype theory, which postulates that experienced translators overrepresent the typical features of the target language because they use certain cognitive links which connect the different related senses of a lexical item more than other links (see Section 2.3.4, 3.6.3 for a

discussion). This may provide an explanation for experienced translators' observed tendency to be less creative on a lexical level, as they rely more on entrenched cognitive links formed by the more extensive use of specific senses compared to others. Translators with less experience are expected to have weaker connections and links, and therefore to be less set in their lexical choices.

Furthermore, despite the fact that creativity is one of the hypothesised components of translation expertise, and is therefore expected to be more visible in the translations of experienced translators compared to inexperienced translators, experienced translators are also expected to be more sensitive to translation norms promoting normalisation. It is possible that the awareness of such norms override translators' creative ability, so that translators sacrifice more creative solutions to avoid the risk of their translations' being rejected. As far as the inexperienced translators are concerned, it may be questioned whether the evidence of lexical creativity observed in this study is really a conscious lexical creativity, or rather a sign of inexperience. It is possible that these inexperienced translators were "inventing" words or borrowing words from the source language, not because they are lexically creative, but because of lexical errors or because they could not think of the correct word to use in their translations. In this sense, inexperienced translators are less sensitive to normalising norms in translation.

In sum, it appears that translations are less inclined to use contractions and are on average less creative than non-translations. In addition, it is evident that inexperienced translators use contractions less than experienced translators, but that they are lexically more creative. It is believed that inexperienced translators use contractions less because of a lack of register sensitivity as they cannot effectively judge in which registers it is acceptable to use contractions. In addition, the higher degree of creative word forms in their target texts is thought to be caused by their lack of experience as they are creating or borrowing words because they could not think of the correct word to use.

4.3.4 Levelling-out

The feature of translated language usually described as "levelling-out" is based on the assumption that translations tend, linguistically, to be inclined towards the middle of a continuum, which essentially means that translators steer a centre course between extremes. Many linguistic features show distinct differences in distribution according to register. If translators do tend to avoid extremes, this theoretically would have the implication that registers in the translational subcorpora would show

less variation for the features investigated than non-translated texts. In addition, it was hypothesised that levelling-out would occur more extensively in the IT subcorpus than in the ET subcorpus as inexperienced translators may be less sensitive to text type or register variation.

However, this study found very little evidence that translation tends to reduce register variation. For all seven features where the interaction between the independent variables corpus and register was investigated, register variation was evident in all three subcorpora, with neither of the translational subcorpora demonstrating a significantly "flattened out" register distribution of features in comparison to the NT subcorpus. Even though in the case of *that*-omission there is some suggestion of levelling-out in the IT subcorpus, there is no statistical support for this observation. This study therefore does not find support for the hypothesis that translations tend to have a relatively consistent middle register, as the translated texts in the corpus, overall, demonstrate sensitivity to register difference.

Significant interaction effects for corpus and register were found for only two features: conjunctive markers and standardised TTR. This interaction effect, however, does not seem to be related to a levelling-out effect. Rather, the IT subcorpus appears to behave differently in certain registers. With the conjunctive markers, the IT subcorpus has a noticeably higher incidence of conjunctive markers in the creative writing and popular writing registers when compared to the other two subcorpora. In the case of standardised TTR, in the IT subcorpus standardised TTR for the creative writing register is much lower when compared to the higher values of the ET and NT subcorpora, and the IT subcorpus also clearly has a higher standardised TTR for the popular writing category as compared to the creative writing register, which is different from the pattern in the other two subcorpora. The interaction effect therefore appears to be limited to registers which are more informal in nature. This suggests that the inexperienced translators' lack of experience or exposure results in them misjudging the appropriateness of the feature for particularly informal registers.

4.4 Conclusion

The first part of this chapter presented the findings of the study, in the form of both descriptive and inferential statistics. These results and findings were then linked to the hypotheses formulated for the different operationalisations (see Section 3.6). Of the eight operationalisations, the results showed statistically significant differences between the three subcorpora for the E1 operationalisation, the S1

and S3 operationalisations, the N1 and N2 operationalisations, and in some features for the L1 operationalisation. The findings for the S2 operationalisation were close to statistical significance. Following this, suggestions were provided as to why a particular hypothesis was supported or rejected, to explain the finding in the context of translation expertise.

This study had two aims. Its secondary aim was to investigate the claim that translations are characterised by their own unique linguistic features and patterns. Against this background, its primary aim was to determine whether the occurrence of these features in the translations of experienced and inexperienced translators can be used to investigate translation expertise. If observable and significant differences in the frequency and distribution of these features were noted in translations produced by these two groups, these differences could then be taken as an indicator of translation expertise.

The findings of this study suggest that certain features of translated language may indeed serve as indicators of translation expertise. Experienced translators are less explicit than inexperienced translators when it comes to *that*-omission, while their texts may be comparatively harder to read than those produced by their inexperienced counterparts. This is because their translations are more lexically diverse, have a comparatively lower readability score, and a longer word length overall. All this indicates a less salient simplification effect in their translations. In addition, experienced translators are more inclined to use contractions, which indicates less of a tendency to normalise, but tend to have a lower incidence of neologisms and loanwords in their translations, which suggests more of a tendency to normalise, than inexperienced translators.

CHAPTER 5: SUMMARY AND CONCLUSION

5.1 Introduction

The objective of this study was threefold. Firstly, it aimed to determine whether there were differences in the frequency and distribution of the features of translated language between translated and non-translated texts. Secondly it aimed to find out if there were noticeable differences between the translations of experienced and inexperienced translators in this regard. Thirdly, it intended to explain how the differences in the frequency and distribution of the features of translated language in the translations of experienced and inexperienced translators can be linked to the concept of translation expertise.

The features of translated language investigated were explicitation, simplification, normalisation and levelling-out. Each of these features was operationalised by means of particular linguistic attributes associated with the feature. In the first instance, it was hypothesised that the translated texts would contain a higher frequency of these features than the non-translated texts, indicating, it was assumed, that the text processing involved in translation differs in distinct ways from the text processing involved in non-mediated text production. In the second instance, it was expected that there would be noticeable differences in the frequency and dispersion of these features in the translations of experienced and inexperienced translators, and that these differences in the third instance may be indicative of translation expertise. In this chapter, a summary of the results of the study is provided and these results are linked to the overall hypothesis that the features of translated language are indicative of translation expertise. Some conclusions are drawn, particularly by placing the findings of the study in the context of existing research on translation expertise. The limitations of the research are outlined, as are some avenues for further research along these lines.

5.2 The features of translated language in translations and non-translations

The results of this study indicated that there are observable and statistically significant differences between translations and non-translations for a number of the features investigated. In terms of **explicitation**, it is evident that translators favour formal completeness in their text production more frequently than non-translators. This is indicated by the fact that translated texts, in this study, had a much lower frequency of omission of the complementiser *that* than non-translated texts. This finding, in other words, can be taken to provide support for the explicitation hypothesis as far as formal

completeness is concerned, since translators were more inclined to use the optional or redundant syntactic element in reporting clauses than non-translators. However, even though the results indicate that explicitation may be a feature of translated language at a formal level, the same does not apply to explicitation on a semantic or propositional level, as reflected in the use of conjunctive markers. For this operationalisation, the differences between the translated and non-translated subcorpora were too small to be statistically significant. These findings possibly suggest that explicitation is not necessarily a conscious process used by translators to make the meaning of their target texts clearer for the target audience. The inclusion of the *that* complementiser might be the consequence of an unconscious process that is caused by a reprocessing of the syntactic structure and finding a corresponding structure in the target language – an interpretation that might be seen as a hypothesis for future testing.

Statistically significant differences were also detected for the operationalisations chosen to study the **simplification** hypothesis. However, these findings only partially support the simplification hypothesis. Simplification did occur in the translations produced by inexperienced translators, as indicated by the fact that this subcorpus had the lowest type-token ratio and highest mean readability score along with the shortest average word length of the three subcorpora. However, the findings for the subcorpus made up of translations done by experienced translators do not support the finding for greater simplification in translated language. Even though the type-token ratio for the subcorpus of work by experienced translators is lower than for the subcorpus of non-translations, the translations done by experienced translators have the lowest readability score and the longest word length of the three subcorpora, meaning that the translations are more difficult to read in these ways than both the non-translated subcorpus and the subcorpus of work by inexperienced translators. This finding, to a degree, does undermine the simplification hypothesis as only one of the translation subcorpora in the corpus was easier to read than non-translations.

For **normalisation**, it was found that both translation subcorpora were less inclined to use contractions, neologisms and loanwords than the non-translated subcorpus. In the case of contractions, there is a tendency among translators to normalise their translations as they conform to a conservative interpretation of a conventional written standard. In the case of neologisms and loanwords, it can be seen that the translations are comparatively more conservative as they are, overall, less creative and more conformist.

For **levelling-out**, analysed by considering register variation for all the linguistic operationalisations used to investigate the other features of translated language, there was very little evidence to support the hypothesis that translators prefer to use a more neutral middle register, thus "levelling out" the differences between registers. The translation subcorpora, in general, show register variation for the distribution of the various features of translated language in much the same way that the subcorpus of originally produced texts does.

From the above, it is evident that there are clear differences between translated and non-translated language in a number of cases, demonstrating that translators tend to use explicitation when it comes to formal completeness, tend to have a narrower vocabulary range, are inclined to raise the formality of their texts and produce texts which are less creative, less marked and more conformist than non-translators. These findings should therefore be considered as providing some support for the overall hypothesis that translated language is qualitatively different from non-translated language – and in some cases the nature of the difference was not as expected.

However, the study was not just limited to the investigation of the features of translated language. The research also aimed to investigate the hypothesis that the features of translated language can be used as indicators of translation expertise; in other words, the hypothesis that these features would demonstrate a significantly different frequency and distribution in the work of experienced and inexperienced translators. The following section reports findings specifically related to this expertise hypothesis.

5.3 The features of translated language and translation expertise

The findings of the study demonstrate that there are significant differences between the translations of experienced and inexperienced translators in a number of cases. In the case of **explicitation**, experienced translators are less explicit in terms of formal completeness than inexperienced translators, as indicated by the fact that they omitted the *that* complementiser more frequently than inexperienced translators, who tend to consistently include it. This provides support for Blum-Kulka's (1986:293) claim that explicitation will be more frequent in the translations of inexperienced translators which she believes is attributable to the fact that the less experienced translators are, the more their interpretation process of the source language will be reflected in the target language. In terms of **simplification**, experienced translators also simplified less frequently than inexperienced

translators, as indicated by their more varied vocabulary, lower scores on the readability index, and longer average word length which, in essence, suggests that their translations are more difficult to read than those produced by inexperienced translators. Mixed findings were found for **normalisation**. Experienced translators used contractions more frequently than inexperienced translators, indicating that they are less inclined to normalise in this sense, but they used neologisms and loanwords less frequently than inexperienced translators, meaning that they normalise more in this regard. As discussed in the previous section, no statistically significant results were found for conjunctive markers as an explicitation operationalisation, or for a reduction in register variation, which would have been indicative of levelling-out.

The above findings suggest that the features of translated language do, in the majority of cases, demonstrate differences in frequency concerning the work of experienced and inexperienced translators. In other words, this study not only provides support for the hypothesis that translated language and non-translated language are qualitatively different along the parameters of the four features investigated, but also that the level of translation expertise influences the frequency and distribution of these features. Viewed from one angle, this means that these features may be used to distinguish the work of experienced and inexperienced translators, and can be taken as indicators of translation expertise. Viewed from another angle, it is very likely that these differences (evident in text **products**) are the consequences of different text-processing and translation strategies used by experienced and inexperienced translators in the translation **process**. The findings therefore also contribute to our understanding of what is meant by translation expertise. Factors which are thought to have an impact on these results include: sensitivity to register requirements, language competence involving morphological, syntactic and vocabulary knowledge, sensitivity to written language conventions, and an awareness of translation norms overall.

The lesser degree of consistent formal explicitation of the experienced translators suggests that they are more aware of register differences than inexperienced translators because they can more effectively judge in which of the different registers it is acceptable to omit the *that* complementiser. They are aware, for instance, that creative and popular writing tend to be more informal in nature and that it is therefore more acceptable to delete the *that* complementiser in these cases. Likewise, in terms of the use of full forms and contractions, experienced translators were also more sensitive to the fact that it is acceptable to use contractions in informal registers whereas full forms are generally

preferred in more formal registers, which again demonstrates their **sensitivity to register requirements**.

The higher type-token ratio in the translations of experienced translators shows that their translations have a greater vocabulary range and, consequently, a greater lexical diversity than those produced by inexperienced translators. This illustrates their experience and their competence as they can draw on a greater range of **morphological, syntactic and vocabulary knowledge** in translation. In addition, the results indicate that the text produced by experienced translators had longer sentence lengths overall, which may indicate that they are more **sensitive to written language conventions**. This is because they may have normalised to a written standard and therefore overrepresented and exaggerated the features that are associated with written language as written language does tend to use longer sentences overall.

The translations of experienced translators were also less creative, less marked, and more conformist than the translations of inexperienced translators, as indicated by the fact that they used neologisms and loanwords less frequently than inexperienced translators overall. This finding was taken to be indicative of professional translators' **awareness of the norms** that have an influence on translation, and their awareness that non-standard choices carry a greater risk of rejection since they may contravene clients' perceptions of acceptable and appropriate language use. As far as the inexperienced translators are concerned, the higher incidence of neologisms and loanwords in their translations may indicate their inexperience and their more limited linguistic resources. It is possible that they could not think of the correct target-language equivalent and therefore resorted to "inventing" words or borrowing from the source text.

All of these strategies may be linked to the initial hypothesis that differences will be evident between translations produced by experienced and inexperienced translators, because experienced translators have a greater awareness of the socio-cultural role and the function of translation in various communication contexts, have a greater range of linguistic resources, and also have better developed cognitive translation skills.

5.4 Implications of the findings, limitations of the study, and future studies

Even though much conceptual and empirical research on translation expertise has been conducted, such research has tended to focus on process-oriented approaches and smaller-scale experimental and cohort studies. This study has aimed to address the lack of systematic and comprehensive corpus-based investigation of translation expertise by utilising the proposed features of translated language as an indicator of translation expertise. One of the contributions made by this study is that quantification of and possibly developing a model concerning translation expertise in terms of the features of translated language, is now possible. This may form the groundwork for others to create a computational model of sorts that could distinguish between the translations of experienced and inexperienced translators. Similar types of studies have been done by Baroni and Bernardini (2005) and Ilisei *et al.* (2010), who have designed computational models which can distinguish non-translations from translations.

However, the study did have its limitations, mainly in terms of corpus composition. The source languages were not represented to the same degree as some, especially Afrikaans, were overrepresented while others, such as Arabic, were underrepresented, which may have had an influence on the results. The different registers in the three subcorpora also could have had a more balanced representation as some were better represented in a particular subcorpus than others.

In addition, even though comparable corpora have the advantage that they are the ideal tool for making hypotheses about norms that operate in a given target culture, they have the disadvantage that they provide little explanatory power and rely on the problematic idea of textual comparability (Bernardini, 2011:11). As such, the quantitative data provided by these methodologies may be misleading when it comes to translation norms or the features of translated language (Bernardini, 2011:11) – an argument which is applicable to this study. In order to overcome this problem, Bernardini (2011:11) suggests that corpus-based studies should combine the comparable-corpus approach with a parallel-corpus methodology, to form bidirectional corpora. These bidirectional corpora, in other words, consist of non-translations and their respective translations, which constitute the parallel corpus component, along with non-translations and comparable translations, which constitute the comparable component (Bernardini, 2007:2). This kind of corpus design will allow translation theorists to check the findings of their comparable corpora against their parallel text components, which will provide more insight into the nature of these features of translated language.

Parallel corpora, in this sense, allow translation researchers to focus on certain aspects or features of translated language for which a comparable corpus cannot provide data. For instance, one could study source-text or -language interference.

In addition, it might also be of benefit to do a study in which different language activities, such as editing or subtitling, performed by individuals who are either experienced or inexperienced in that particular field are compared to one another to determine whether similar results are found or whether the occurrence of these features is limited to translation specifically. Such a study could address the concerns of Chesterman (2004:10), who questions whether these features of translated language are not necessarily specific to translation per se, but are caused by "extra-constrained communication" in general. This notion of extra-constrained communication means that these features will also be noticeable in other communication contexts that pose extra difficulties, such as writing or speaking in a second language.

5.5 Conclusion

The aim of this study was to address the lack of systematic and comprehensive corpus-based analyses of translation expertise by using the hypothesised features of translation as indicators of translation expertise in order to acquire more descriptive information about the phenomenon. As such, the study set out to determine what the differences are in the occurrence and distribution of the features of translated language in the translations of experienced and inexperienced translators, by means of a corpus-based methodology. It was found that there are observable differences in the frequency and distribution of these features associated with different levels of expertise, as well as between translations and non-translations. It was argued that the differences between the translations produced by experienced and inexperienced translators may be indicative of differences in the translation strategies they used, and these differences are the consequence of underlying differences in expertise. These underlying differences related to translation expertise are believed to include register sensitivity, language competence that involves syntactic, morphological and vocabulary knowledge, an awareness of written language conventions, and sensitivity to translation norms.

REFERENCES

ADAB, B. 2000. Evaluating translation competence. (*In* Schäffner, C. & Adab, B., *eds*. *Developing translation competence.* Amsterdam: John Benjamins. p. 215-228.)

ALBIR, A.H. & ALVES, F. 2009. Translation as a cognitive activity. (*In* Munday, J., *ed. The Routledge companion to translation studies*. London: Routledge. p. 54-73.)

ALVES, F. 2005. Bridging the gap between declarative and procedural knowledge in the training of translators: meta-reflection under scrutiny. *Meta,* 50(4):1-17.

ALVES, F. & GONÇALVES, J.L. 2003. A Relevance Theory approach to the investigation of inferential processes in translation. (*In* Alves, F., *ed. Triangulating translation*. Amsterdam: John Benjamins. p. 3-24.)

ALVES, F. & GONÇALVES, J.L. 2007. Modelling translator's competence: relevance and expertise under scrutiny. (*In* Gambier, Y., Shlesinger, M. Stolze, R., *eds*. *Doubts and directions in translation studies*. Amsterdam: John Benjamins. p. 41-55.)

ALVES, F. & GONÇALVES, J.L. 2010. Relevance and translation. (*In* Gambier, Y. & Van Doorslaer, L., *eds*. *Handbook of translation studies*. Amsterdam: John Benjamins. p. 279-284.)

ALVES, F., PAGANO, A., NEUMANN, S., STEINER, E. & HANSEN-SCHIRRA, S. 2010. Translation units and grammatical shifts: towards an integration of product and process-based translation research. (*In* Shreve, G.M. & Angelone, E., *eds*. *Translation and cognition*. Amsterdam: John Benjamins. p. 109-141.)

ALVES, F. & VALE, D. 2011. On drafting and revision in translation: a corpus linguistics oriented analysis of translation process data. *TC3: Compution, Corpora, Cognition,* 1(1):105-122.

ANDERSON, J.R. 1982. Acquisition of cognitive skill. *Psychological Review*, 89(1):369-406.

ANDERSON, J.R. 1983. *The architecture of cognition*. Cambridge: Harvard University Press. 345 p.

ANGELELLI, C.V. 2009. Using a rubric to assess translation ability: defining the construct. (*In* Angelelli, C.V. & Jacobson, H.E., *eds*. *Testing and assessment in translation and interpreting studies*. Amsterdam: John Benjamins. p. 13-48.)

ANGELONE, E. 2010. Uncertainty, uncertainty management, and metacognitive problem solving in the translation task. (*In* Shreve, G.M. & Angelone, E., *eds*. *Translation and cognition*. Amsterdam: John Benjamins. p. 17-40.)

ARMSTRONG, S., CATFREY, C., FLANAGAN, M., KENNY, D., O'HAGAN, M. & WAY, A. 2006. Leading by example: automatic translation of subtitles via EBMT. *Perspectives: Studies in Translatology*, 14(3):163-184.

BAER, J.B. & BYSTROVA-MCINTYRE, T.B. 2009. Assessing cohesion: developing assessment tools on the basis of comparable corpora. (*In* Angelelli, C.V. & Jacobson, H.E., *eds*. *Testing and assessment in translation and interpreting studies*. Amsterdam: John Benjamins, p. 159-183.)

BAKER, M. 1995. Corpora in translation studies: an overview and some suggestions for future research. *Target*, 7(2):223-243.

BAKER, M. 1996. Corpus-based translation studies: the challenges that lie ahead. (*In* Somers, H.L., *ed. Terminology, LSP and translation: studies in language engineering in honour of Juan C. Sager*. Amsterdam: John Benjamins, p. 175-186.)

BAKER, M. 2007. Patterns of idiomaticity in translated vs. non-translated text. *Belgian Journal of Linguistics*, 21:11-21.

BARONI, M. & BERNARDINI, S. 2006. A new approach to the study of translationese: machine-learning the difference between original and translated text. *Literary and Linguistic Computing*, 21(3):259-274.

BAUMGARTEN, N., MEYER, B. & ÖZÇETIN, D. 2008. Explicitness in translation and interpreting: a critical review and some empirical evidence (of an elusive concept). *Across Languages and Cultures*, 9(2):177-203.

BIBER, D. 1995. *Dimensions of register variation: a cross-linguistic comparison*. Cambridge: Cambridge University Press. 428 p.

BIBER, D., CONRAD, S. & REPPEN, R. 1998. *Corpus linguistics: investigating language structure and use*. Cambridge: Cambridge University Press. 300 p.

BECHER, V. 2010a. Abandoning the notion of "translation-inherent" explicitation: against a dogma of translation studies. *Across Languages and Cultures*, 11(1):1-28.

BECHER, V. 2010b. Towards a more rigorous treatment of the explicitation hypothesis in translation studies. *Trans-kom*, 3(1):1-25.

BECHER, V. 2011. When and why do translators add connectives: a corpus-based study. *Target*, 23(1):26-47.

BEEBY, A. 2000. Evaluating the development of translation competence. (*In* Schäffner, C. & Adab, B., *eds*. *Developing translation competence*. Amsterdam: John Benjamins. p. 185-198.)

BEEBY, A. 2004. Language learning for translators: designing a syllabus. (*In* Malmkjær, K., *ed. Translation in undergraduate degree programs*. Amsterdam: John Benjamins. p. 39-65.)

BERNARDINI, S. & ZANETTIN, F. 2004. When is a universal not a universal? Some limits of current corpus-based methodologies for the investigation of translation universals. (*In* Mauranen, A. & Kujamäki, P., *eds*. *Translation universals: do they exist?* Amsterdam: John Benjamins. p. 51-62.)

BLUM-KULKA, S. 1986. Shifts of cohesion and coherence in translation. (*In* Venuti, L., *ed.* 2004. *The translation studies reader*. 2nd ed. London: Routledge. p. 290-305.)

BLUM-KULKA, S. & LEVENSON, E.A. 1984. Universals of lexical simplification. (*In* Færch, C. & Kasper, G., *eds*. *Strategies in interlanguage communication*. London: Longman. p. 119-139.)

CARL, M. 2009. Triangulating product and process data: quantifying alignment units with keystroke data. *Copenhagen Studies in Language*, 38:225-247.

CASTAGNOLI, S. 2008. *Regularities and variations in learner translations: a corpus-based study of conjunctive explicitation*. Italy: University of Pisa. (Thesis – Ph.D.) 165 p.

CHESTERMAN, A. 1997. *Memes of translation: the spread of ideas in translation theory*. Amsterdam: John Benjamins. 219 p.

CHESTERMAN, A. 2000. Teaching strategies for emancipatory translation. (*In* Schäffner, C. & Adab, B., *eds*. *Developing translation competence*. Amsterdam: John Benjamins. p. 77-90.)

CHESTERMAN, A. 2004a. Beyond the particular. (*In* Mauranen, A. & Kujamäki, P., *eds*. *Translation universals: do they exist?* Amsterdam: John Benjamins. p. 34-49.)

CHESTERMAN, A. 2004b. What is a unique item? (*In* Gambier, Y., Shlesinger, M. & Radegundis, S., *eds*. *Doubts and directions in translation studies*. Amsterdam: John Benjamins. p. 3-13.)

CHESTERMAN, A. 2006. Interpreting the meaning of translation. (*In* Suominen, M., *ed. A man of measure. Festschrift in honour of Fred Karlsson on his 60th birthday*. Turku: Linguistic Association of Finland. p. 3-11.)

CHESTERMAN, A. 2010. Why study translation universals? *Acta Translatologica Helsingiensia*, 1:38-48.

CHESTERMAN, A. 2011. Translation universals. (*In* Gambier, Y. & Van Doorslaer, L., *eds. Handbook of translation studies*. Amsterdam: John Benjamins. p. 175-179.)

CINTRÃO, H.P. 2011. Development of translation competence in novices: a corpus design and key-logging analysis. (*In* O'Brien, S., *ed. Continuum studies in translation: cognitive explorations of translation*. London: Continuum. p. 86-107.)

COLLINS COBUILD. 1990. *Collins Cobuild English Grammar Grammar*. London: Collins. 438 p.

CRYSTAL, D. 2003. *The Cambridge encyclopaedia of the English language*. 2nd ed. New York: Cambridge University Press. 506 p.

DELAERE, I., DE SUTTER, G. & PLEVOETS, K. 2012. Is translated language more standardized than non-translated language? Using profile-based correspondence analysis for measuring linguistic distances between language varieties. *Target*, 24(2): 203-224.

ENGLUND DIMITROVA, B. 2003. Explicitation in Russian–Swedish translations: sociolinguistic and pragmatic aspects. (*In* Englund Dimitrova, B. & Pereswetoff-Morath, A., *eds. Swedish contributions to the thirteenth international congress of Slavists, Ljubljana*. Lund: Lund University Press. p. 21-31.)

ENGLUND DIMITROVA, B. 2005. *Expertise and explicitation in the translation process*. Amsterdam: John Benjamins. 303 p.

ESKOLA, S. 2004. Untypical frequencies in translated language: a corpus-based study on a literary corpus of translated and non-translated Finnish. (*In* Mauranen, A. & Kujamäki, P., *eds. Translation universals: do they exist?* Amsterdam: John Benjamins. p. 83-99.)

EVEN-ZOHAR, I. 1978. The position of translated literature within the literary polysystem. (*In* Venuti, L., *ed.* 2004. *The translation studies reader*. 2nd ed. London: Routledge. p. 199-204.)

FISCHER, R. 1998. *Lexical change in present-day English: a corpus-based study of the motivation, institutionalization and productivity of creative neologisms*. Regensburg: Tübingen. 198 p.

FRASER, J. 2000. The broader view: how freelance translators define translation competence. (*In* Schäffner, C. & Adab, B., eds. *Developing translation competence*. Amsterdam: John Benjamins. p. 51-62.)

FRAWLEY, W. 1984. Prolegomenon to a theory of translation. (*In* Venuti, L., *ed.* 2000. *The translation studies reader*. London: Routledge. p. 250-263.)

GALE, W. & CHURCH, K.W. 1993. A program for aligning sentences in bilingual corpora. *Computational Linguistics*, 19(1):75-102.

GARANT, M. 2008. *Current trends in translation teaching and learning*. Helsinki: University of Helsinki Department of Translation Studies. 301 p.

GILE, D. 2009. *Basic concepts and models for interpreter and translator training*. Amsterdam: John Benjamins. 283 p.

GÖPFERICH, S. 2009. Towards a model of translation competence and its acquisition: the longitudinal study 'TransComp'. (*In* Göpferich, S. Jakobsen, A.L. & Mees, I.M., *eds*. *Behind the mind: methods, models and results in translation process research*. Copenhagen: Samfundslitteratur. p. 11-37.)

GÖPFERICH, S. 2013. Translation competence: explaining development and stagnation from a dynamic systems perspective. *Target*, 25(1):61-76.

GÖPFERICH, S. & JÄÄSKELÄINEN, R. 2009. Process research into the development of translation competence: where are we, and where do we need to go? *Across Languages and Cultures*, 10(2):169-191.

GÖPFERICH, S., BAYER-HOHENWARTER, G.B., PRASSL, F. & STADLOBER, J. 2011. Exploring translation competence acquisition: criteria of analysis put to the test. (*In* O'Brien, S., *ed. Continuum studies in translation: cognitive explorations of translation*. London: Continuum. p. 57-85.)

GREENFIELD, P. & SUBRAHMANVAM, K. 2003. Online discourse in a teen chatroom: new codes and new modes of coherence in a visual medium. *Applied Developmental Psychology*, 24:713-738.

GUTT, E.-A. 2000. Issues of translation research in the inferential paradigm of communication. (*In* Olohan, M., *ed. Intercultural faultiness – research models in translation studies 1: textual and cognitive aspects*. Manchester: St. Jerome. p. 161-179.)

GUTT, E.-A. 2004. Challenges of metarepresentation to translation competence. (*In* Fleischmann, E., Schmitt, P.A. & Wotjak, G., *eds*. *Translationskompetenz. Tagungsberichte der LICTRA*. Stauffenberg: Tübingen. p. 77-89.)

HALLIDAY, M.A.K. & MATTHIESSEN, M.I.M. 2004. *An introduction to functional grammar*. 3rd ed. London: Arnold. 689 p.

HARRIS, B. 2010. Unprofessional translation. Blog: http://unprofessionaltranslation.blogspot.com Date of access: 19 Oct. 2010.

HALVERSON, H. 1998. Translation studies and representative corpora: establishing links between translation corpora, theoretical/descriptive categories and a conception of the object of study. *Meta*, 43(4):494-514.

HALVERSON, S. 2003. The cognitive basis of translation universals. *Target*, 15(2):197-241.

HALVERSON, S. 2008. Psycholinguistics and cognitive approaches. (*In* Baker, M. & Saldanha, G., *eds*. *Routledge encyclopaedia of translation studies*. 2nd ed. London: Routledge. p. 211-216.)

HILL, T. & LEWICKI, P. 2006. *Statistics: methods and applications*. Cape Town: StatSoft. 832 p.

HOLMES, J. 1972. The name and nature of translation studies. (*In* Venuti, L., *ed.* 2004. *The translation studies reader*. 2nd ed. London: Routledge. p. 180-192.)

ICE (International Corpus of English). 2012. International Corpus of English: homepage. http://ice-corpora.net/ice/index.htm Date of access: 11 Nov. 2010.

ILISEI, I., DIANA, I., PASTOR, G.C. & RUSSLAN, M. 2010. Identification of translationese: a machine learning approach. (*In* Gelbukh, A., *ed. Computational linguistics and intelligent text processing lecture notes in computer science*. Berlin: Heidelberg. p. 503-511.)

JÄÄSKELÄINEN, R. 1993. Investigating translation strategies. (*In* Tirkkonen-Condit, S. & Laffling, J., *eds*. *Recent trends in empirical translation research*. Joensuu: University of Joensuu. p. 99-116.)

JÄÄSKELÄINEN, R. 1996. Hard work will bear beautiful fruit: a comparison of two think-aloud protocol studies. *Meta*, 41(1):60-74.

JÄÄSKELÄINEN, R. 1999. *Tapping the process: an explorative study on the cognitive and affective factors involved in translating.* University of Joensuu publications in the Humanities 22. Joensuu: University of Joensuu.

JÄÄSKELÄINEN, R. 2002. Think-aloud protocol studies into translation: an annotated bibliography. *Target*, 14(1):107-136.

JÄÄSKELÄINEN, R. 2008. Think-aloud protocols. (*In* Baker, M. & Saldanha, G., *eds*. *Routledge encyclopaedia of translation studies*. 2nd ed. London: Routledge. p. 290-293.)

JÄÄSKELÄINEN, R. 2010. Are all professionals experts? Definitions of expertise and reinterpretation of research evidence in process studies. (*In* Shreve, G.M. & Angelone, E., *eds*. *Translation and cognition*. Amsterdam: John Benjamins. p. 213-227.)

JIMÉNEZ-CRESPO, M. 2011. The future of general tendencies in translation: explicitation in web localisation. *Target*, 23(1):3-25.

KAMENICKÁ, R. 2007. Defining explicitation in translation. *BRNO Studies in English*, 33(1):45-57.

KELLY, D. 2005. *A handbook for translator trainers: a guide to reflective practice*. Manchester: St. Jerome. 173 p.

KENNY, D. 1998. Creatures of habit? What translators usually do with words. *Meta*, 43(4):515-523.

KENNY, D. 2001. *Lexis and creativity in translation: a corpus-based study*. Manchester: St Jerome. 254 p.

KENNY, D. 2008. Corpora. (*In* Baker, M. & Saldanha, G., *eds*. *Routledge encyclopaedia of translation studies*. 2nd ed. London: Routledge. p. 59-62.)

KIRALY, D. 1995. *Pathways to translation: pedagogy and process*. Kent, OH: Kent State University Press. 179 p.

KIRALY, D. 2000. *A social constructivist approach to translator education: empowerment from theory to practice*. Manchester: St. Jerome. 207 p.

KLAUDY, K. 2008. Explicitation. (*In* Baker, M. & Saldanha, G., *eds*. *Routledge encyclopaedia of translation studies*. 2nd ed. London: Routledge. p. 104-108.)

KRUGER, A. 2002. Corpus-based translation research: its development and implication for general, literary and Bible translation. *Acta Theologica Supplementum*, 2:70-106.

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

KUJAMÄKI, P. 2004. What happens to "unique items" in learners' translations? (*In* Mauranen, A. & Kujamäki, P., *eds*. *Translation universals – do they exist?* Amsterdam: John Benjamins. p. 187-204).

KÜNZLI, A. 2004. Risk taking: trainee translators vs professional translators – a case study. *The Journal of Specialised Translation*, 2:34-49.

LAVIOSA, S. 1998a. The English comparable corpus: a resource and a methodology. (*In* Bowker, L., Cronin, M. & Kenny, D., *eds*. *Unity in diversity: current trends in translation studies*. Manchester: St Jerome. p. 101-112.)

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

LAVIOSA, S. 2002. *Corpus-based translation studies: theory, findings, applications*. New York: Rodopi. 138 p.

LAVIOSA, S. 2003. Corpora and translation studies. (*In* Granger, S., Lerot, J. & Petch-Tyson, S., *eds*. *Corpus-based approaches to contrastive linguistics and translation studies*. New York: Rodopi. p. 45-54.)

LAVIOSA, S. 2008a. Universals. (*In* Baker, M. & Saldana, G., *eds. Routledge encyclopaedia of translation studies*. 2nd ed. London: Routledge. p. 306-310.)

LAVIOSA, S. 2008b. Description in the translation classroom: universals as a case in point. (*In* Pym, A., Shlesinger, M. & Simeoni, D., *eds*. *Beyond descriptive translation studies: investigations in homage to Gideon Toury*. Amsterdam: John Benjamins. p. 119-132).

LEESON, J., SAEED, J., LEONARD, C., MACDUFF, A. & BYRNE-DUNNE, D. 2006. Moving heads and moving hands: developing a digital corpus of Irish sign language. *IT & T Annual Conference Proceedings 2006*. Carlow: Ireland. p. 33-43.

LÖRSCHER, W. 1996. A psycholinguistic analysis of translation processes. *Meta*, 1:26-32.

LÖRSCHER, W. 2005. The translation process: methods and problems of its investigation. *Meta*, 50(2):597-608.

LÖRSCHER, W. 2009. Form- and sense-oriented approaches to translation revisited. *Translation Quarterly*, 51:1-28.

LOUW, B. 1993. Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. (*In* Baker, M., Francis, G. & Togini-Bonelli, E., *eds*. *Text and technology*. Amsterdam: John Benjamins. p. 157-176.)

MACKENZIE, R. 2004. The competencies required by the translator's roles as a professional. (*In* Malmkjær, K., *ed. Translation in undergraduate degree programs*. Amsterdam: John Benjamins. p. 31-38.)

MALMKJÆR, K. 1992. The relevance of translation studies. *Ilha do Destierro*, 28:25-36.

MALMKJÆR, K. 1998. Love thy neighbour: will parallel corpora endear linguists to translators? *Meta*, 43(4):534-541.

MALMKJÆR, K. 2008. Translation competence and the aesthetic attitude. (*In* Pym, A., Shlesinger, M. & Simeoni, D., *eds*. *Beyond descriptive translation studies: investigations in homage to Gideon Toury*. Amsterdam: John Benjamins. p. 293-309.)

MARTÍN, R.M. 2000. Translation strategies: somewhere over the rainbow. (*In* Beeby, A., Ensinger, D. & Presas, M., *eds*. *Investigating translation*. Amsterdam: John Benjamins. p. 129-138.)

MARTÍN, R.M. 2009. Expertise and environment in translation. *Mutatis Mutandis: Revista Latinoamericana de Traducción*, 2(1):24-37.

MARTÍN, R.M. 2010. Leave no stone unturned – on the development of cognitive translatology. *Translation and Interpreting Studies*, 5(2):145-162.

MASSEY, G. 2005. Process-oriented translator training and the challenge for e-learning. *Meta*, 50(2):626-633.

MAURANEN, A. 2004. Corpora, universals and interference. (*In* Mauranen, A. & Kujamäki, P., *eds*. *Translation universals – do they exist?* Amsterdam: John Benjamins. p. 65-82.)

MAURANEN, A. 2008. Universal tendencies in translation. (*In* Anderman, G.M. & Rogers, M., *eds*. *Incorporating corpora: the linguist and the translator*. Ontaro: Multilingual Matters. p. 39-46.)

MAY, R. 1997. Sensible elocution: how translation works in and upon punctuation. *The Translator*, 3(1):1-20.

MOE, Z. 2010. Register shifts in translations of popular fiction from English into Slovene. (*In* Gile, D., Hansen, G. & Pokorn, N.K., *eds*. *Why translation studies matter*. Amsterdam: John Benjamins. p. 125-136.)

MOLINA, L. & ALBIR, A.H. 2002. Translation techniques revisited: a dynamic and functionalist approach. *Meta*, 47(4):498-512.

MONDAHL, M. & JENSEN, K.A. 1996. Lexical search strategies in translation. *Meta*, 41(1):97-113.

MUNDAY, J. 1998. A computer-assisted approach to the analysis of translation shifts. *Meta*, 43(4):1-16.

MUNDAY, J. 2009. *The Routledge companion to translation studies*. London: Routledge. 287 p.

MUNDAY, J. 2011. Looming large: a cross-linguistic analysis of sematic prosodies in comparable reference corpora. (*In* Kruger, A., Wallmach, K. & Munday, J., *eds*. *Corpus-based translation studies: research and applications*. London: Continuum. p. 169-186).

MUTESAYIRE, M. 2004. Apposition markers and explicitation: a corpus-based study. *Language Matters: Studies in the Languages of Africa*, 35(1):54-69.

NEUBERT, A. 2000. Competence in language and translation. (*In* Schäffner, C. & Adab, B., *eds*. *Developing translation competence*. Amsterdam: John Benjamins. p. 3-18.)

NELSON, G., WALLIS, S. & AARTS, B. 2002. *Exploring natural language: working with the British component of the International Corpus of English.* Amsterdam: John Benjamins. 351 p.

NIDA, E.A. 1964. *Toward a science of translating, with special reference to principles and procedures involved in Bible translating*. Leiden: Brill. 335 p.

NORD, C. 2005. *Text analysis in translation: theory, methodology, and didactic application of a model for translation*. New York: Rodopi. 274 p.

O'BRIEN, S. 2013. The borrowers: researching the cognitive aspects of translation. *Target,* 25(1):5-17.

OLOHAN, M. 2001. Spelling out the optionals in translation: a corpus based study. (*In* Rayson, P., Wilson, A., McEnery, T., Hardie, A. & Khoja, S., *eds. Proceedings of the corpus linguistics 2001 conference, UCREL Technical Papers*. Lancaster: Lancaster University Press. p. 423-432.)

OLOHAN, M. 2002. Comparable corpora in translation research: overview of recent analyses using the translational English corpus. (*In* Yuste, E., *ed. LREC Language resources in translation work and research workshop proceedings*. Paris: ELRA. p. 5-9.)

OLOHAN, M. 2004. *Introducing corpora in translation studies*. London: Routledge. 220 p.

OLOHAN, M. & BAKER, M. 2000. Reporting *that* in translated English: evidence for subconscious processes of explicitation? *Across Languages and Cultures,* 1:142-172.

OROZCO, M. 2000. Building a measuring instrument for the acquisition of translation competence in trainee translators. (*In* Schäffner, C. & Adab, B., *eds*. *Developing translation competence*. Amsterdam: John Benjamins. p. 199-215.)

ØVERÅS, L. 1998. In search of the third code: an investigation of norms in literary translation. *Meta,* 43(4):557-570.

OXFORD DICTIONARIES ONLINE. 2012. Oxford dictionaries: homepage. http://oxforddictionaries.com Date of access: 8 Aug. 2010.

OXFORD UNIVERSITY PRESS. 1997. Genstat 5: reference manual. Oxford: Oxford University Press. 777 p.

PACTE, 2000. Acquiring translation competence: hypotheses and methodological problems of a research project. (*In* Beeby, A., Ensinger, D. & Presas, M., *eds*. *Investigating translation*. Amsterdam: John Benjamins. p. 99-105.)

PACTE. 2002. Exploratory tests in a study of translation competence. *Conference Interpretation and Translation*, 4(2):41-69.

PACTE, 2005. Investigating translation competence: conceptual and methodological issues. *Meta*, 50(2):609-619.

PACTE, 2008. First results of a translation competence experiment: knowledge of translation and efficacy of the translation process. (*In* Kearns, J., *ed. Translator and interpreter training: issues, methods and debates*. London: Continuum. p. 104-126.)

PACTE, 2009. Results of the validation of the PACTE translation competence model: acceptability and decision making. *Across Languages and Cultures*, 10(2):207-230.

PACTE, 2011. Results of the validation of the PACTE translation competence model: translation project and dynamic translation index. (*In* O'Brien, S., *ed. Continuum studies in translation: cognitive explorations of translation*. London: Continuum. p. 30-56.)

PÁPAI, V. 2004. Explicitation: a universal of translated text? (*In* Mauranen, A. & Kujamäki, P., *eds*. *Translation universals: do they exist?* Amsterdam: John Benjamins. p. 143-165.)

PASTOR, G.C., MITKOV, R., AFZAL, N. & PEKAR, V. 2008. A corpus-based NLP study of convergence and simplification. Available: http://www.mt-archice.info/AMTA-2008-Corpas Date of access: 19 Oct. 2010.

PUURTINEN, T. 1998. Syntax, readability and ideology in children's literature. *Meta*, 43(4):524-533.

PUURTINEN, T. 2004. Explicitation of clausal relations: a corpus-based analysis of clause connectives in translated and non-translated Finnish children's literature. (*In* Mauranen, A. & Kujamäki, P., *eds*. *Translation universals – do they exist?* Amsterdam: John Benjamins. p. 165-177).

PYM, A. 2003. Redefining translation competence in an electronic age: in defence of a minimalist approach. *Meta*, 48(4):481-497.

PYM, A. 2005. Explaining explicitation. (*In* Károly, K. & Fóris, A., *eds. New trends in translation studies – in honour of Kinga Klaudy*. Budapest: Akadémiai Kiadó. p. 29-34.) (Draft version at http://www.tinet.org/~apym/on-line/translation/explicitation_web.pdf, unpaginated.)

PYM, A. 2008. On Toury's laws of how translators translate. (*In* Pym, A., Shlesinger, M. & Simeoni, D., *eds*. *Beyond descriptive translation studies – investigations in homage to Gideon Toury*. Amsterdam: John Benjamins. p. 311-328.)

PYM, A. 2010. Translation research terms – a tentative glossary for moments of perplexity and dispute. Available: http://www.tinet.cat/~apym/on-line/research _methods/2010_terms Date of access: 19 Oct. 2010.

RISKU, H., DICKINSON, A. & PIRCHER, R. 2010. Knowledge in translation studies and translation practice: intellectual capital in modern society. (*In* Gile, D., Hansen, G. & Pokorn, N.K., *eds*. *Why translation studies matters*. Amsterdam: John Benjamins. p. 83-94.)

ROBINSON, D. 2007. *Becoming a translator: an introduction to the theory and practice of translation.* 2nd ed. London: Routledge. 295 p.

ROSNER, M. 2004. Corpus-driven bilingual lexicon extraction. (*In* Pace, G. & Cordina, J., *eds*. *Proceedings of Computer Science Annual Research Workshop organised by University of Malta*. p. 78-85).

SALDANHA, G. 2008. Explicitation revisited: bringing the reader into the picture. *Trans-kom*, 20:20-35.

SÉGUINOT, C. 1988. Pragmatics and the explicitation hypothesis. *Meta*, 1(2):106-113.

SÉGUINOT, C. 2000. Management issues in the translation process. (*In* Tirkkonen-Condit, S. & Jääskeläinen, R., *eds*. *Tapping and mapping the processes of translation and interpreting*. Amsterdam: John Benjamins. p. 143-148.)

SCOTT, M. 1996, 2004, 2008. WordSmith Tools [versions 3.0, 4.0, 5.0]. Oxford: Oxford University Press.

SCARPA, F. 2006. Corpus-based quality-assessment of specialist translation: a study using parallel and comparable corpora in English and Italian. (*In* Gotti, M. & Šarčević, S., *eds*. *Insights into specialised translation*. Frankfurt: Peter Lang. p. 155-172.)

SHREVE, G. 2002. Knowing translation: cognitive and experiential aspects of translation expertise from the perspective of expertise studies. (*In* Riccardi, A., *ed. Translation studies: perspectives on an emerging discipline*. Cambridge: Cambridge University Press. p. 150-171.)

SHREVE, G. & ANGELONE, E. 2010. Translation and cognition: recent developments. (*In* Shreve, G. & Angelone, E., *eds. Translation and cognition*. Amsterdam: John Benjamins. p. 1-13.)

SHULKA, S. & CONNOR-LINTON, J. 2006. Language change. (*In* Fasold, R. & Connor-Linton, J., *eds. An introduction to language and linguistics*. Cambridge: Cambridge University Press. p. 275-305.)

SNELL TRAMPUS, R. Aspects of a theory of norms and some issues in teaching translation. (*In* Riccardi, A., *ed. Translation studies: perspectives on an emerging discipline*. Cambridge: Cambridge University Press. p. 38-55.)

TANSKANEN, S.K. 2006. *Collaborating towards coherence: lexical cohesion in English discourse*. Amersterdam: John Benjamins. 208 p.

TEMPLE, B. & YOUNG, A. 2004. Qualitative research and translation dilemmas. *Qualitative Research*, 4(2):161-178.

TIRKKONEN-CONDIT, S. 2002. Translationese – a myth or an empirical fact? *Target*, 14(2):207-220.

TIRKKONEN-CONDIT, S. 2004. Unique items – over- or under-represented in translated language? (*In* Mauranen, A. & Kujamäki, P., *eds*. *Translation universals – do they exist?* Amsterdam: John Benjamins. p. 176-184).

TIRKKONEN-CONDIT, S. 2005. The monitor model revisited: evidence from process research. *Meta*, 50(2):405-414.

TOURY, G. 2004. Probabilistic explanations in translation studies: welcome as they are, would they qualify as universals? (*In* Mauranen, A. & Kujamäki, P., *eds*. *Translation universals – do they exist?* Amsterdam: John Benjamins. p. 15-32).

TOURY, G. 2012. *Descriptive translation studies and beyond*. 2nd ed. Amsterdam: John Benjamins. 350 p.

TYMOCZKO, M. 1998. Computerised corpora and the future of translation studies. *Meta*, 43(4):652-660.

ULRYCH, M. 2005. Training translators: programmes, curricula, practices. (*In* Tennent, M., *ed. Training for the new millennium*. p. 3-35.)

VALENTINI, C. 2006. A multimedia database for the training of audiovisual translators. *The Journal of Specialised Translation*, 6:68-83.

VAN ROOY, B. & ESTERHUIZEN, I. 2011. Die gebruik van konjunksiemerkers by nagraadse studente en gepubliseerde navorsing in die Afrikaanse taal- en letterkunde. *SAALT Journal for Language Teaching*, 45(1):67-86.

VENUTI, L. 2000. Translation, history, narrative. *Meta*, 50(3):800-816.

VINAY, J.P. & DARBELNET, J. 1958. *Comparative stylistics of French and English: a methodology for translation.* Translated from the French by J.C. Sager and M.-J. Hamel. Amsterdam: John Benjamins. 358 p.

WILLIAMS, D. 2005. Recurrent features of translation in Canada: a corpus-based study. Ottowa: University of Ottawa. (Dissertation – Ph.D.) 340 p.

WILLIAMS, J. & CHESTERMAN, A. 2001. *The map: a beginner's guide to doing research in translation studies.* Manchester: St. Jerome. 149 p.

YAJUN & ZAIXIN. 2008. Translated English and universals of translation. *English Today*, 24(4):26-33.

YUAN, Y. & GAO, F. 2008. Universals of translation: a corpus-based investigation of Chinese translated fiction. *Proceedings of The International Symposium on using Corpora in Contrastive and Translation Studies.* p. 1-25.

ZANETTIN, F. 2011. Hardwiring corpus-based translation studies: corpus encoding. (*In* Kruger, A., Wallmach, K. & Munday, J., *eds*. *Corpus-based translation studies: research and applications*. London: Continuum. p. 103-123).

ZANETTIN, F. 2012. *Translation-driven corpora*. Manchester: St. Jerome. 244 p.