Levelling-out and register variation in the translations of experienced and inexperienced translators: a corpus-based study[1]

Karien Redelinghuys

School of Languages, North-West University, Vaal Triangle Campus, South Africa
E-mail: karien.redelinghuys@nwu.ac.za

Abstract
Explicitation, simplification, normalisation and levelling-out, the four features of translation proposed by Baker (1996), have attracted considerable attention in translation studies. Although the first three have been studied extensively, levelling-out has been the subject of less empirical investigation. Furthermore, there are no studies to date that have investigated the extent to which levelling-out occurs in translations by experienced translators and inexperienced translators. In this study, levelling-out is operationalised in terms of register. It is hypothesised that less register variation will be apparent in translations by inexperienced translators and, in keeping with the features of translation hypothesis, it is predicted that select linguistic features will demonstrate less register variation in translations than in non-translations. A custom-built corpus was compiled to test these hypotheses. While some light is shed on how translation expertise contributes to register sensitivity and the distribution of certain features across different registers, little evidence could be found for levelling-out as register variation is evident in the translation corpora.

Keywords: translation expertise, levelling-out, register variation, universals of translation

1.   Introduction

Research based on translation expertise, which is also sometimes referred to as translation competence, has been a growing area of investigation in translation studies since the 1990s (Lesznyák 2008: 31; Martín 2014: 2). These studies have not only focused on how translation expertise may be conceptualised and defined, but also on how this expertise is acquired by translators and how knowledge thereof could be applied to translators' training (Albir, Alves, Dimitrova and Lacruz 2015: 5). While a variety of factors contribute to this increasing interest, the fact that academic programmes for translator training started to develop in earnest during this time no doubt played an important role (Lesznyák 2008: 31). The development of translation training programmes of necessity presupposes a definition and demarcation of the

---

[1] This article is based on a component of a dissertation entitled *Using the features of translated language to investigate translation expertise: A corpus-based study* submitted in fulfilment of the requirements for the degree Master of Arts in Language Practice at the North-West University.

necessary skills and knowledge that constitute translation expertise. For this reason, research aimed at delimiting the competence became a necessity for the purposes of training. During the same time period, process-oriented research on translation also expanded significantly. Typical data-collection methods in the process-oriented paradigm include keylogging, eye-tracking, think-aloud protocols, and, more recently, neuro-imaging, all of which in different ways serve as a direct record of the translation process, and an indirect indication of some of the cognitive processing involved. While these kinds of data are often combined with product analysis, process-oriented translation research is distinct from other kinds of translation research (i.e. some types of corpus-based and descriptive research) that attempts to infer translation processes from translation products.

As pointed out by O'Brien (2015: 5), the relationship between process-oriented translation research and research on translation expertise is so close that developments in the one field have mutually fed into the other. Presently, process-oriented research on translation expertise is a burgeoning area of investigation in descriptive translation studies (see O'Brien 2015 for an overview), with research groups and individual researchers alike dedicated to investigating various dimensions of translation expertise and its acquisition (cf. e.g. PACTE 2009, 2011a, 2011b; Angelone 2010; Ehrensberger-Dow and Massey 2014; Schwieter and Ferreira 2014; Göpferich 2015). The emphasis on process-oriented, experimental methodologies in research on translation expertise has meant that product-oriented methodologies have been used comparatively infrequently. Where product-oriented methodologies are used, they tend to use translation product analysis on a smaller scale (e.g. the analysis of translation products produced by participants in a particular process-oriented study of translation expertise), and larger-scale, quantitative product-based methodologies, specifically corpus approaches, are an under-represented area of enquiry, regardless of the calls of translation researchers such as Alves, Pagano, Neumann, Steiner and Hansen-Schirra (2010) and Alves and Vale (2011) for the incorporation of corpora in process-oriented translation studies. Therefore, currently the overlap is limited between translation expertise research and yet another flourishing area in translation studies, namely corpus-based translation studies – a product-based methodology used to study a number of aspects that are related to translated language (Saldanha and O'Brien 2013). The investigation on which this article reports is based on the assumption that a better understanding of and new insights into translation expertise can be gained by using the corpus-based approach to test, complement and extend the results from process-oriented studies.

A primary concern of corpus-based translation studies is the study of claim that translations demonstrate distinctive linguistic patterns or regularities that differ from those of non-translations in what has been referred to as the features (or universals) of translation (see Zanettin 2012, 2013). Following Baker (1996), the features are often categorised as simplification, explicitation, normalisation and levelling-out. Explicitation refers to a tendency to spell things out in translation rather than leaving them implicit; simplification refers to an inclination to simplify the message and language of translations to "make things easier for the reader"; normalisation is a tendency to exaggerate target-language features and to conform to the typical patterns of the target language; and levelling-out occurs when the degree of similarity between translated texts is measurably greater than non-translations, which are more dispersed and less homogenous (Baker 1996: 18–1 184). Further potential candidates proposed as possible features of translation include source-language and source-text transfer phenomena expressed in Toury's (2012) law of interference and the underrepresentation of unique target language features (Tirkkonen-Condit 2002; see also Eskola 2004). Research into the features of translated language has been

prolific and considerable progress has been made towards the empirical validation of some features of translated language in the two decades since its proposal (see Zanettin 2013 for an overview), so that some consensus is developing regarding the textual features that distinguish translations from non-translations (Redelinghuys and Kruger 2015: 296).

However, the occurrence of these features, is a point of contention and a number of aspects associated with the original proposal of translation universals have been subject to theoretical scrutiny by scholars in the field. A critical point concerns the conceptualisation of the features of translation in the sense that it does not make 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 translator, ideologies, period, language or genre (Tymoczko 1998: 653; Steiner 2012: 4). In other words, the relation between norms and translation phenomena may be subject to historical variation – a fact which is not accounted for within the paradigm of translation universals. A different point of criticism is that the way in which universal claims are formulated and investigated is sometimes unclear, which makes replication of previous studies difficult or in some cases even impossible (Chesterman 2011: 178).

Despite some of the objections to the features of translated language, the study of their occurrence is generally considered to be beneficial to translation studies because investigating these phenomena helps translation scholars and translators to gain a better understanding of the nature of translation (Chesterman 2011: 178). In this respect, one of the main reasons for investigating these features of translation is because it has the potential to raise translators' awareness of the unconscious or conscious processing involved in translation (Chesterman 2011: 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).

While there is considerable corpus-based research on the features of translated language, this research has mostly been limited to corpora of translations carried out by professional and experienced translators, which suggests these features could occur as a result of translation strategies and/or language processing related to translation expertise. Laviosa (2008: 307) believes that the features of translated language are the result of three factors, namely the constrained cognitive processing that occurs in the translation process, translation's communicative role and translators' awareness of their sociocultural roles and positions. It is possible that these factors play different roles for inexperienced and experienced translators. If this is the case, it is possible that the features of translation will be manifested to different degrees in translations produced by experienced and inexperienced translators. Therefore, these features may act as indicators of translation expertise. In this regard, the hypothesis of Redelinghuys and Kruger's (2015) study was based on the assumption that if the features of translated language are conceptualised as the textual "sediment" of translation strategies and/or processes, differences in inexperienced and experienced translators' translation strategies and/or linguistic processing ought to be manifested in terms of different frequencies of these features in their translation output. Their findings provide substantial (though not unqualified) support for the hypothesis in terms of explication, simplification and normalisation.

While Redelinghuys and Kruger's (2015) study illustrates the relationship between translation expertise and the occurrence of the features of translated language, they do not discuss the effect of experience on levelling-out, the fourth universal proposed by Baker (1996). Following Kruger and Van Rooy (2012), levelling-out will be investigated as a feature across register in this study; that is to say levelling-out is conceptualised as a reduction of the distinctness of various registers in favour of a more neutral "middle" register. Register will arguably present a fruitful field of enquiry when it is considered that inexperienced and experienced translators tend to differ from one another with regard to their sensitivity of 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, Conrad and Reppen 1998: 137). In this sense it has to be kept in mind that register awareness is considered to be one of the key elements of translation expertise (cf. e.g. Kelly 2005; Nord 2005; Angelelli 2009; Colina 2015; Olohan 2016) even though it is not a well-documented topic in studies of translation expertise (Colina 2015: 211). As such, it can be hypothesised that translations by experienced and inexperienced translators of different registers will be homogenous to different degrees, which in turn, should give an indication of their different levels of experience.

While this study assumes that levelling-out will be manifested to different degrees in translations by experienced and inexperienced translators, it should be pointed out levelling-out is possibly one of the more problematic features of translated language. Its problematic nature is partly due to the fact that it has been subject to limited systematic investigation in comparison with the other features of translation (Williams 2005: 45; Pym 2007: 175; Zanettin 2012: 20; Xiao and Dai 2014: 20) and the research conducted on levelling-out to date has yielded inconclusive results. These results may in part have been the result of a conceptual problem as there appears to an assortment of understandings of what levelling-out entails. Considering the limited amount of attention levelling-out has received, there is consequently an additional need to qualify it as a feature of translated texts and to investigate the claim that translated texts are less varied and more homogenised than non-translated texts.

This study is an extension of the one carried out by Redelinghuys and Kruger (2015) in that it uses the same features of translation as proposed indicators of translation expertise in respect of explicitation, simplification and normalisation. This investigation starts with an overview of ways in which levelling-out has been conceptualised in other studies, its various operationalisations and the results of other investigations. It then discusses the hypotheses formulated for this study, as well as the corpus composition and the methodology used. The results of the various operationalisations are discussed and interpreted in terms of levelling-out.

2.   Conceptualisations of levelling-out

The hypothesised universals of simplification, normalisation, explicitation and levelling-out together are indicative of a process whereby translation as a particular kind of writing activity consists of the 'de-complexification' of language (Zanettin 2012: 13). It is thought that translators simplify the source text meaning by conforming to typical language patterns and by cutting on the peripheries of more creative language use, while adding supplementary information where necessary (Zanettin 2012: 13). This de-complexification process may result in the degree of similarity between groups of translations being measurably greater than the similarity of non-translations, as translators are expected to steer a middle course between extremes (Baker 1996: 184). Baker (1996: 184) refers to this proposed translational feature as

levelling-out, which she defines as translations' inclination to "gravitate towards the centre of a continuum". Laviosa (2002: 72) prefers the term "convergence" to describe "the relatively higher level of homogeneity of translated texts".

While Pym (2008: 318) does not argue against the notion that translations have a distinctive profile, he finds it difficult to distinguish between the features of explicitation, normalisation, and simplification as these three features overlap to a large degree. He consequently argues against levelling-out as it follows that any extreme explicitation, normalisation and simplification would result in the gravitation of translations toward the centre of a continuum. However, Xiao and Dai (2014: 20) are of the opinion that it is justifiable to argue that explicitation, simplification and normalisation do not have to occur as extremes for them to be considered as translational features. They occur relative to the source language or to the target language and it is the occurrence of these features that make translations more convergent and homogenous (Xiao and Dai 2014: 20).

The notion that levelling-out occurs on a cline, caught somewhere between the tension of the source language and the target language, is echoed by Hansen-Schirra and Steiner (2012), who are of the opinion that levelling-out can be located on a norm continuum of translation properties. At the one end of their proposed continuum is "shining-through" (which is when source-language norms are met) and at the other side is normalisation (or when target-language norms are met). Levelling-out, along with bleaching and hybridisation, may be found anywhere between the patterns typical of the source-language and those of the target-language (Hansen-Schirra and Steiner 2012: 272).

Another conceptualisation of levelling-out is that it can occur as an occurrence or feature across register (Kruger and Van Rooy 2012). Register can be seen as a frequency-based phenomenon: characteristic linguistic features of a situational context come to be characteristic because of their reoccurrence in that context, leading to their conventionalisation (Neumann 2014: 36). As such, register has come to be recognised as a central issue in the distribution of linguistic features (Neumann 2014: 36), in that the dispersion of linguistic features can be expected to vary across registers (Biber 2014: 7). However, according to the levelling-out hypothesis translations will differ from non-translated texts in that they will be "more similar to each other in terms of some (set of) linguistic features [as the] range of variation among translations are assumed to be smaller than for otherwise similar original texts" regardless of register or language (Steiner 2012: 4). In other words, based on the assumption that translations gravitate towards the centre of a continuum and are more homogenous than non-translated texts, the register variability typical of non-translations may be reduced in favour of a neutral middle register, leading to levelling-out. This middle register may be evident in the distribution of specific linguistic features across different registers.

As pointed out by Grabowski (2013: 275), invalidating or confirming the levelling-out hypothesis depends on its operationalisation, which is complicated by the fact that there is seemingly no consensus on this issue. As will be seen from the discussion of previous studies in section 3, levelling-out has been operationalised in different ways, which includes operationalisations based on lexical density, type-token ratio, readability scores and, to a lesser extent, register variability. None of the existing studies have pertinently included expertise as a factor in investigating the levelling-out hypothesis.

3.   Levelling-out: operationalisations and results

Laviosa (1998a) does not explicitly attempt to study levelling-out, but finds some evidence for its occurrence in one of her simplification studies. In her corpus-based study of non-translated English newspaper articles and comparable English translations from different source languages, she finds that translations are more homogenous in terms of their lexical density. Lexical density is conceptualised as the ratio of the number of lexical words to the number of running words in a particular text (Laviosa 1998a: 104). In a study of a comparable corpus of translational and non-translational English narrative prose; however, the results do not confirm greater homogeneity in terms of lexical density (Laviosa 1998b). She consequently postulates that levelling-out possibly only applies to registers other than narrative prose (Laviosa 1998b: 565), but has not pursued this line of enquiry in subsequent work.

Scarpa (2006) studies a variety of operationalisations to investigate explicitation, simplification and normalisation in a parallel corpus of specialist English texts and their Italian translations. Scarpa's (2006) corpus consists of different specialist text types (information technology, humanities and social sciences) written in different registers (textbooks, low-level popular science, academic articles). For lexical density (one of her operationalisations for simplification), the results indicate that translated texts are more homogenous than their source texts, which leads her to conclude that she found evidence for levelling-out. The fact that Scarpa (2006) finds some support for levelling-out in lexical density in specialist registers provides some support for Laviosa's (1998b) proposal that levelling-out may be applicable only to registers other than narrative prose or fiction.

The results of Neumann's (2012) study also indicates that levelling-out may be influenced by register. In a bidirectional translation corpus of English and German texts, mood options are studied as an indicator of social role relationship in two registers. Mood options are an aspect of context of situation that is related to the relationship between addressee and sender and is, as such, indicative of level of authority (Neumann 2012: 198). One of the registers in Neumann's (2012) study consists of contemporary literary texts, while the other consists of letters in the name of or from the chief executive officer of different companies to their shareholders, which are informative and persuasive in nature. For the translations of the business letters, the results show that, while the original texts show a particular range of variation, the translated texts do not change the frequency of the various mood options in relation to the reference corpora, which is taken as indicative of levelling-out. The frequency of the different mood options for the literary translations, by contrast, are systematically similar to the source texts, which is taken as indicative of shining-through of the source texts.

Based on the studies of Laviosa (1998a; 1998b), Scarpa (2006) and Neumann (2012), it could be easy to surmise that levelling-out is characteristic of more formal rather than informal registers. However, some studies have found evidence for levelling-out in informal registers or little evidence for levelling-out in more formal registers. Grabowski's (2013) investigation, for example, finds evidence for levelling-out in a custom-built corpus of contemporary non-translated and translated literary Polish texts. Grabowski (2013) uses a methodology based on Principal Components Analysis and Cluster Analysis (two methods that have been used widely in stylometry and authorship attribution) to test the levelling-out hypothesis based on word frequency and word distribution. The results of these two analyses confirm the levelling-out hypothesis that translated texts are more similar than non-translations.

Yuan and Gao (2008) use type-token ratio, sentence length and lexical density to study levelling-out, as well as simplification in a comparable corpus of non-translated and translated Chinese fiction. Readability scores are also used to determine if levelling-out occurs. They hypothesise that the translations in their corpus will have more homogenous scores, which would be expressed as lower standard deviation, than the fiction non-translations. The results of their study only provide partial support for levelling-out as the variables of standardised type-token ratio, sentence length and lexical density confirm levelling-out, but the readability scores are not less varied for the translations in comparison with the non-translations.

Another study that uses readability scores is Williams's (2005) investigation based on a comparable corpus of English and French translations and non-translations on a range of non-literary subjects. She hypothesises levelling-out will occur if translations produce readability scores with narrower ranges of highs and lows (along a continuum that ranges from "very easy" to read "to very difficult") than non-translations, as measured by standard deviation. However, the results fail to support the levelling-out hypothesis consistently. Even though the standard deviation for the English translations is comparatively lower, the results are not statistically significant. In addition, the French translations' scores are comparatively higher than those of the non-translations.

Whereas the studies discussed mainly focus on levelling-out in one register (or two in the case of Neumann's (2012) study), Xiao's (2011) corpus consists of different registers, which includes news, nonfiction, academic, and fiction writing. His study is based on a comparable corpus of original Chinese, translational Chinese and British English along with an English-Chinese parallel corpus. Attention is paid to word clusters and reformulation markers in all the registers combined (rather than separately). It is found that more frequent use is made of word clusters in translational Chinese than in non-translational Chinese, which is thought to be an effect of English source language influence. It is also found that reformulation markers are used as an explicitation strategy so as to make translational Chinese texts stylistically and orally simpler than non-translations. The findings lead Xiao (2011) to conclude that the qualitative and quantitative differences for word clusters and reformulation markers over the entire corpus provide some evidence for levelling-out.

The study of Kruger and Van Rooy (2012) is the first that specifically focuses on levelling-out in terms of register variability. In their study, levelling-out is operationalised as a smaller degree of register variability in a translated subcorpus than in a non-translated subcorpus, occurring as a consequence of the effects of translation. Their study is based on a comparable corpus of original and translated English texts produced in South Africa, and they use a variety of operationalisations to study explicitation, normalisation, simplification and levelling-out in terms of register variability. These operationalisations include the use of the optional complementiser *that*, frequency of full forms as opposed to contracted forms, frequency of linking adverbials, frequency of loanwords and coinages, frequency of lexical bundles, lexical diversity, and mean word length. They do not find overall support for their hypothesis as register differences are noted for all the linguistic features investigated irrespective of translated or non-translated status, but some subtle differences between the subcorpora are noted nonetheless. It is found, for instance, that the popular writing register of the translated corpus does not show the same degree of informality since the non-translated corpus is more closely aligned with written-language norms. This finding leads Kruger and Van Rooy (2012: 62) to conclude that the features typical of formal registers may only become visible in less formal

registers or, in other words, that levelling-out occurs because the informal registers are made more formal. They point out one shortcoming of their study, namely that most of the translated texts were translated from Afrikaans and suggest that a translation corpus with multiple source-languages be used to minimise interference effects.

The levelling-out studies discussed in this section focus on texts produced by experienced translators, but considerably less attention has been paid to translations by inexperienced translators. The study by Pastor, Mitkov, Afzal and Pekar (2008) is an exception. In their study, Pastor et al. (2008) investigate levelling-out in a comparable corpus composed of subcorpora of medical and technical translations into Spanish and a collection of Spanish non-translated texts. The subcorpus of medical translations is divided into two divisions: translations by professional translators and translations by student translators. In their study, they operationalise the levelling-out hypothesis in terms of lexical density, lexical richness, sentence length, use of discourse markers, proportion of simple sentences, and types of syntactic constructions used. Only discourse markers show the translated subcorpora to be more similar to one another, which provides some evidence for levelling-out (Pastor et al. 2008). However, they point out that the occurrence of the discourse markers in the medical translations by experienced and inexperienced translators are more varied when compared with each other. The authors do not provide any suggestions as to what may have contributed to this finding nor do they discuss the implications of this finding in terms of translation expertise. This finding, nonetheless, suggests that levelling-out may have an expertise dimension associated with its occurrence and that it may be worthwhile to study the degree to which translations by experienced translators and translations by inexperienced translators are more homogenous or less homogenous when compared with each other.

4.   Research questions and hypotheses

From the literature review it is evident that not only has little attention been paid to levelling-out from a translation expertise perspective, but there also has been limited investigation of levelling-out in terms of register variation. Two research questions are formulated for this particular investigation, namely: (1) To what degree does levelling-out occur in translations by experienced and inexperienced translators? (2) And is there validity to the claim that register differences tend to be levelled out in translation as opposed to non-translations?

Predictions regarding the occurrence of levelling-out are difficult to make not only because levelling-out has been the subject of few empirical studies, but also because it has not been conceptualised from an expertise dimension. On the one hand, it has to be kept in mind that experienced translators will be more closely aligned with a professional standard and will be more consistently subject to the conventions and norms of this standard. This may encourage decomplexification of the source language for the target audience, consequently resulting in higher instances of explicitation, simplification and normalisation, causing translations that are more homogenous in terms of the distribution of particular linguistic features across different registers. On the other hand, their experience also encourages a greater awareness of text type, register sensitivity and audience awareness, which could result in a greater degree of variation as they adjust their translations to ensure their target texts are acceptable to the target audience.

For the inexperienced translators, it is postulated that less variation of specific linguistic features will be evident in their translations as they are generally less aware of register preferences

(Fawcett 1997: 83). This lack of register sensitivity is illustrated by Deeb's (2005: 245) study, which found that even though translation students show a degree of register appreciation in interviews, there are inconsistencies in their use of register markers. Students, for instance, tend to replace specialist lexical items with common terms and are inclined to use colloquial vocabulary in formal texts (Fawcett 1997: 83). Their translations may also 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. Against this background, the first hypothesis proposes that levelling-out will occur to a greater degree in the work of inexperienced translators than experienced translators.

In the second instance, it is hypothesised that if translations tend to gravitate towards the centre of a continuum, as is assumed by the proposed feature of levelling-out, the register variability characteristic of non-translated language will be reduced in favour of a more neutral middle register in translations. In other words, it is hypothesised that the register effect will be a less strong for the linguistic features investigated in the translations than in comparison with non-translations. If this is the case, it may be assumed that translation is the cause of the levelling-out of the variation.

5.   Corpus compilation

This study uses a custom-built comparable corpus of English texts that consists of three subcorpora: translations by inexperienced translators (the IT subcorpus), translations by experienced translators (the ET subcorpus), and non-translations (the NT subcorpus). The translations chosen for the IT subcorpus were done by either student translators or laypersons – translators who have no experience in professional translation. The student translations were collected from several universities in Poland and South Africa and were produced by students who were enrolled for translation programmes. Texts were selected from the Internet to represent translations produced by laypersons. In all these cases the layperson specified that the text is a translation. They also clearly indicated that they were not professional translators or that they had little translation experience, but produced the translations for enjoyment or with the intention of getting feedback from native English speakers.

The texts contained in the ET and NT subcorpora were collected from the Internet, from printed published works or from existing corpora. In terms of the ET subcorpus, some texts were selected from an existing translating corpus (see Kruger and Van Rooy 2012). In general, it was assumed that these texts are instances of authentic, naturally occurring translations that 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. In the case of the NT subcorpus, texts were taken from the International Corpus of English for South Africa (ICE-SA) (see ICE 2012) and from the British National Corpus (BNC) to match the representation of British English and South African English in the translation corpora. Seeing as translations done in a variety of contexts is included in the translation subcorpora, it was considered essential to include British, American and South African English in the NT subcorpus. British English accounts for 45% of the NT subcorpus, followed by South African English with 32% and American English with 23%.

All of the texts in the corpus were produced from 1982 to 2012 and the texts were either samples or full texts of no less than 1 000 words. The corpus is spread over 85 translations in the IT subcorpus, 87 translations in the ET subcorpus and 84 texts in the NT subcorpus.

6.   Register and source languages

For the study, registers were classified using some of the written text categories of the ICE-corpora. The registers are: academic, creative, instructional, popular, and reportage writing. The final token count for the five registers in the three subcorpora is indicated in table 1.

Table 1: Number of tokens by register in the three subcorpora

	IT	ET	NT
Academic writing	32 310	33 504	33 939
Creative writing	30 462	29 840	33 018
Instructional writing	34 039	26 133	25 928
Popular writing	40 381	37 482	36 768
Reportage	26 448	28 818	28 910
Total	**163 640**	**155 777**	**158 563**

Although the token count of table 1 suggests the corpus is relatively balanced, the corpus was not completely balanced in terms of the different source languages in the different registers. The register imbalance is in part due to the genre preferences of lay translators who prefer to translate creative texts. The token count and the percentage of the different source languages' representation in the IT and ET subcorpora are presented in table 2.

Table 2: Representation of source languages for the two translation subcorpora

	IT		ET	
		Percentage		Percentage
Language	Tokens	of corpus	Tokens	of corpus
Germanic				
Afrikaans	56 067	34.26	56 234	36.11
Danish	1 003	0.61	5 996	3.85
Dutch	4 188	2.56	11 007	7.07
German	6 631	4.05	6 391	4.10
Norwegian	5 838	3.57	5 298	3.40
Swedish	11 077	6.77	6 224	4.00
Romance				
French	12 854	7.86	17 728	11.38
Italian	1 751	1.07	1 610	1.03
Spanish	15 206	9.29	9 258	5.94
Altaic				
Japanese	12 357	7.55	6 138	3.94
Korean	1 176	0.72	2 336	1.50
Slavonic				
Polish	14 377	8.79	3 499	2.25
Russian	5 413	3.31	3 898	2.52
Bulgarian	1 013	0.62	0.00	0.00

Language	IT		ET	
	Tokens	Percentage of corpus	Tokens	Percentage of corpus
Sino-Tibetan				
Chinese	13 384	8.18	13 249	8.51
Thai	0.00	0.00	1 393	0.89
Finno-Ugric				
Estonian	0.00	0.00	1 501	0.96
Indic				
Nepali	0.00	0.00	2 854	1.83
Semitic				
Arabic	1 305	0.80	1 128	0.72

As shown in table 2, the translation subcorpora have a strong representation of Germanic source languages – especially Afrikaans. This overrepresentation is due to the fact that the majority of the student translations have Afrikaans as a source language, which had to be reflected in the ET subcorpus for reasons of comparability.

The NT subcorpus consists mainly of examples of British and South African English with some American English. British and American English were included because they are global standard varieties and South African English was included to match its representation in the two translation subcorpora. Other Englishes were excluded to limit sources of variability in the NT subcorpus. The overall number of tokens and percentage of the three varieties in the NT subcorpus in terms of the five different registers are shown in table 3.

Table 3: Representation of English varieties in the NT subcorpus for the five registers

Register	American English		British English		South African English	
	Tokens	Percentage of corpus	Tokens	Percentage of corpus	Tokens	Percentage of corpus
Academic writing	5 211	15.35	15 741	46.38	12 987	38.27
Creative writing	6 562	19.87	20 072	60.79	6 384	19.33
Instructional writing	6 384	24.62	11 127	42.91	8 417	32.46
Popular writing	11 159	30.35	13 479	36.66	12 130	32.99
Reportage	7 080	24.49	11 773	40.72	10 057	34.79
Total	**36 396**	**22.94**	**72 192**	**45.49**	**49 975**	**31.57**

7.   Data collection and data processing

Data were collected and extracted using the WordList and Concord functions in WordSmith 5.0 corpus-analysis software (Scott 2008). *Statistica 11* (Statsoft, 2012) was used for statistical processing. Factorial Analysis of Variance (ANOVA) was used to determine if there is interaction between the independent variables of register and corpus for each of the dependent variables,

which would indicate if the three subcorpora demonstrate different register-related preferences for the specific linguistic feature studied.

8.   Linguistic features studied

A selection of linguistic features were investigated in terms of their frequency in the five registers specified to determine whether the dispersion of these linguistic features across the registers is more homogenous in the translated subcorpora when compared with the non-translated subcorpus, and most undifferentiated for register in the IT subcorpus. The features selected for the investigation are: omission of the complementiser *that*, conjunctive markers, standardised type-token ratio, word length, readability scores, contractions, and neologisms. These particular linguistic features were chosen as they can be used as operationalisations of the other features of translated language, in particular explicitation, simplification and normalisation (see Redelinghuys and Kruger (2015) for a discussion on the selection of the features). This study is similar to Kruger and Van Rooy (2012) in its conceptualisation of levelling-out in terms of register variability but it should be noted that translations from different source languages were included in the present investigation (rather than just translations from Afrikaans), and some of the linguistic features differ from those investigated by their study.

9.   Findings

9.1   Omission of the complementiser *that*

The complementiser *that* is considered to be a redundant syntactic element, because its retention is optional with reporting verbs. According to Biber, Johansson, Leech, Conrad and Finegan (1999: 680) even though retention or omission of the *that* complementiser has no effect on meaning, there are discourse factors that will influence whether it is used or not. The *that* complementiser is almost always retained in academic writing (Biber and Conrad 2001: 180), which is thought to be because academic prose is characterised by careful production circumstances with an informational, expository purpose (Biber et al. 1999: 680). It is also the norm for the *that* complementiser to be retained in newspaper reports because of the construction types that characterise this type of writing (such as the use of a passive in the main clause) (Biber et al. 1999: 683). In more informal registers, such as fiction and conversation, the complementiser *that* is more frequently omitted. This is thought to be the case because these registers are characterised by involved and interpersonal purposes (Biber and Conrad 2001: 180). The complementiser *that*, in other words, is typically omitted more frequently in informal registers (such as those of creative writing and popular writing) than in formal registers.

Studying omission of the *that* complementiser was selected based on the assumption that its omission (as opposed to its inclusion) signifies a lower degree of explicitness. If it is omitted to a lower degree in translated corpora as opposed to non-translations it would support the explicitation hypothesis. For this investigation, the reporting verbs that control complement clauses listed by the *Cobuild English Grammar Dictionary* (Bullon, Krishnarmurthy, Manning and Todd 1990) were used as search terms (see Addendum A). The incidence of *that*-omission as a ratio of the possible occurrences where a choice may have been made between the retention or the omission thereof was calculated and analysed.

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. As is evident in figure 1, 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. Therefore, the two translated subcorpora 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 some visual indication of a levelling-out effect in translation, which is more noticeable in the IT subcorpus, but statistical support for this effect is lacking.
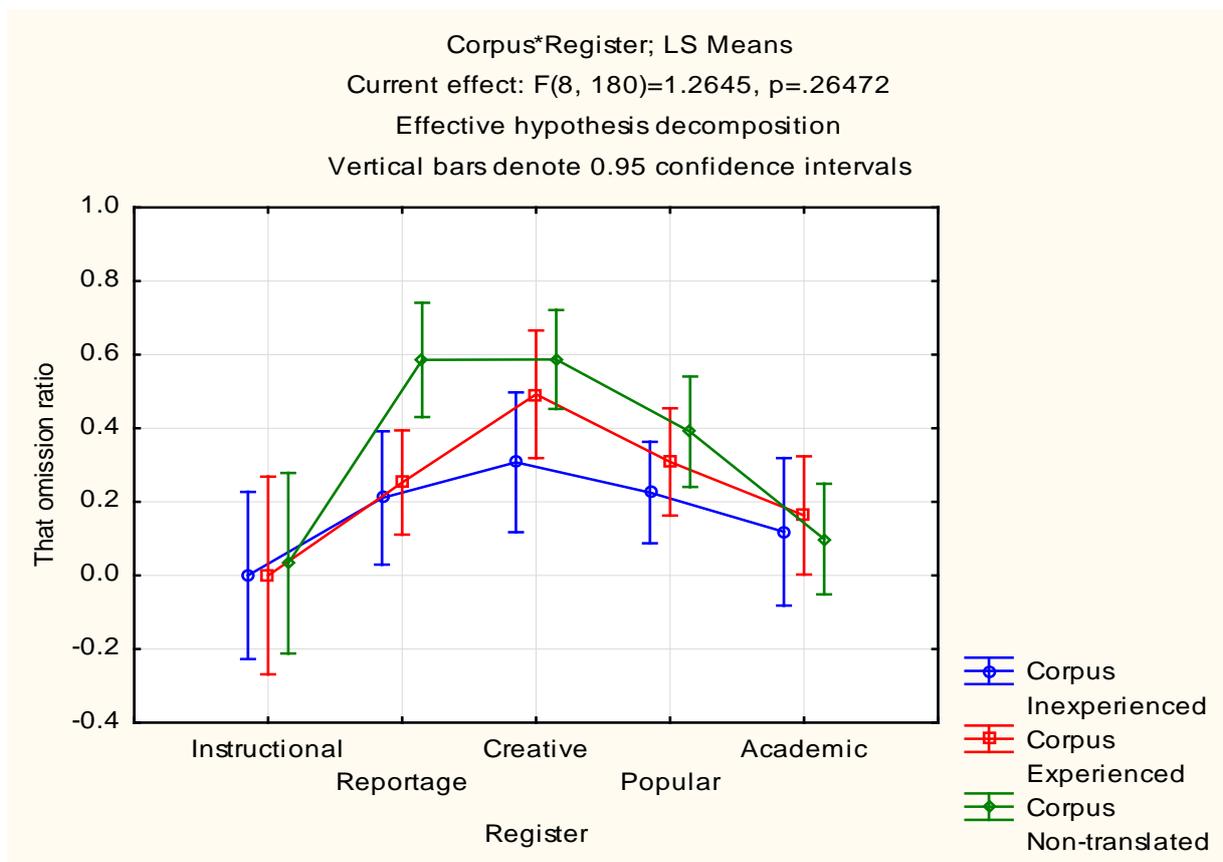
Figure 1: *That*-omission ratio in the three corpora and five registers

9.2   Conjunctive markers

Conjunctive markers can be considered as indicators of explicitation as they explicitly mark propositional relationships between units of discourse. Conjunctive relations include the specification of a reference, repetition of information provided previously so as to avoid ambiguity, the expansion of condensed passages, the addition of explanatory phrases, and the addition of cohesive devices to promote text flow. For this investigation, the use of the conjunctive markers of enhancement, extension and elaboration (Halliday and Matthiessen 2004: 541) was analysed (see Addendum B). The word *and*, which is a conjunctive marker of

extension, was not included in the study as it would have required an unmanageable amount of manual sorting because of its ambiguous nature as a clause coordinator and phrase coordinator (for explicit treatment of this issue, refer to Van Rooy and Esterhuizen, 2011).

Nesi and Basturkmen (2009: 24) liken conjunctions to what Biber et al. (1999) refer to as linking adverbials. It has been found that linking adverbials are used more commonly in academic writing where the emphasis is on building arguments and conveying logical coherence as they allow writers to mark their arguments' development by relating one proposition to another (Biber et al. 1999: 767). They are used less commonly in newspaper reports and fiction as readers are expected to infer cause-effect relationships (Biber and Conrad 2009: 119).

The factorial ANOVA demonstrates a statistically significant interaction effect for corpus and register ($F(8, 239)=3.11$, $p<0.05$) (see figure 2). In the reportage register of the IT subcorpus, conjunctive markers occur slightly more as in the same register in the NT and ET subcorpora. In the popular writing and creative writing registers, the IT subcorpus has a higher incidence of conjunctive markers as opposed to the academic and instructional registers, where it has the lowest incidence. This finding is somewhat contrary to expectations. Typically, one would expect conjunctives to be more frequent in registers where it is important to mark the ideas between ideas visibly, such as academic writing, and less in registers where cause and result are expected to be inferred from chronological sequences. However, inexperienced translators evidently misjudge the register expectations by overtly linking ideas in registers that are intrinsically more informal in nature than those that are more formal, which suggests inexperienced translators only explicitate by using conjunctive markers in the more informal registers. It may well be that they consciously or unconsciously overcompensate in the translation process due to their inexperience.

Overall, the ET and NT subcorpora use conjunctive markers in a similar way over the five registers, even though the ET subcorpus is inclined to use them slightly more in the reportage, creative writing and academic writing registers. Therefore, a similar pattern is 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 that in the NT subcorpus in terms of this feature.

In sum, the varied pattern of the IT subcorpus and the similarity between the ET and NT subcorpora suggests that translation-related levelling-out in the use of conjunctive markers across the five registers in the translated subcorpora does not occur. However, it appears that the inexperience of translators plays a role with the IT subcorpus demonstrating a significantly different pattern for the distribution of this feature compared with the other two subcorpora. Inexperienced translators appear to misjudge the appropriateness of the addition of conjunctive markers particularly in more informal registers.
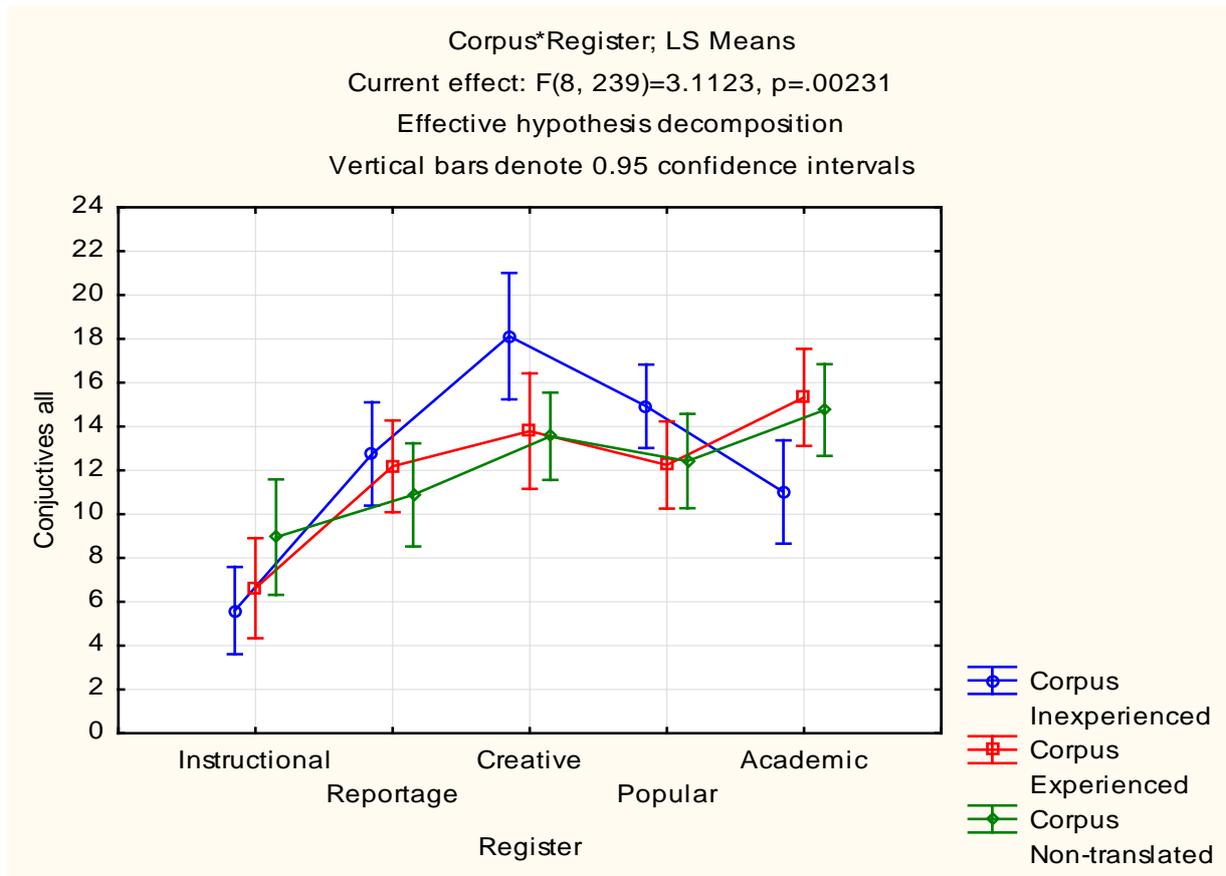
**Figure 2:** Conjunctive markers in the three subcorpora and five registers

9.3   Standardised type-token ratio

Type-token ratio (TTR) refers to the ratio of unique word forms to the running words in a text and is used as a measurement of lexical diversity (Teich 2003: 21). The higher the TTR, the less repetitive and more varied the vocabulary, while the lower the ratio the more repetitive and extensive the vocabulary of a text. Fiction tends to have a relatively high TTR since the focus is on the elegance and form of expression (Biber et al. 1999: 54). Reportage also tends to have a high TTR (which is comparatively lower than that of fiction) due to the density of nominal elements of this register (Biber et al. 1999: 54). Technical-writing text types, such as academic writing, tend to have a lower TTR frequency than other text types as technical writing derives its exactness or preciseness from the repeated use of a word with a technically defined meaning (Biber and Finegan 1994: 342). In these registers, alternative expressions are undesirable in the discussion of technical subjects as readers may try to infer minor differences in meaning (Biber and Finegan 1994: 342).

In research on the features of translated language, TTR is typically used as an indicator of simplification at the lexical level. This is based on the assumption that translations are characterised by more repetition and a higher ratio of grammatical words that consequently results in a simplified lexicon and a more limited lexical range, which will be reflected in a lower TTR (Zanettin 2012: 15). As the text lengths for this corpus were varied, the type-token ratio was recalculated for every 1 000 words in the text before the average for the entire text was computed.

There is a statistically significant interaction effect ($F_{(8, 234)}=3.57$, $p<0.001$) for the variables corpus and register (see figure 3). From figure 3 it can be seen that, in keeping with the simplification hypothesis, the average standardised TTR for the NT subcorpus is generally much higher than those of the translated subcorpora, which means that the non-translations are characterised by a much more varied vocabulary with the exclusion of the instructional register. Interestingly enough, the reportage, creative writing and popular writing registers in the NT subcorpus have similar TTRs – a pattern mirrored by the ET subcorpus. This pattern is indicative of experienced translators' sensitivity to lexical diversity across various registers even though their translations' TTR values do not match those of the non-translations. However, it is questionable whether a case such as this would constitute levelling-out. As pointed out section 2, levelling-out is a feature that is supposed to be unique to translation. The fact that these three registers in the translations of experienced translators mirror the flattened-out distribution of non-translations arguably does not constitute levelling-out. In this case, it is rather a register-specific effect independent of translation status. Allowance also to be made for the fact that the instructional writing and academic writing registers have much lower TTRs for both these subcorpora, which means that the findings do not provide support for the levelling-out hypothesis.

The IT subcorpus shows a flattened-out TTR distribution for the reportage, creative writing and academic writing registers. However, it is evident that the popular writing register is noticeably higher when compared with the other registers in the subcorpus, while the instructional writing corpus has a lower TTR. The fact that all five of the registers do not show a relatively consistent middle does therefore not support the levelling-out hypothesis. It is interesting to note creative writing's TTR in comparison with the other registers in the IT corpus. While it would be expected that creative writing would have the most varied vocabulary, the inexperienced translators' vocabulary range for this register is similar to those of reportage and academic writing. Even though it may be easy to surmise that the lower TTR displayed by creative writing is as a result of lower language proficiency, which is reflected in shorter and simpler words, the TTR for popular writing is noticeably higher, which shows that they have the ability to produce texts with a greater vocabulary range. Therefore, it is possible that they misjudge the register conventions of creative writing with those of more formal registers.
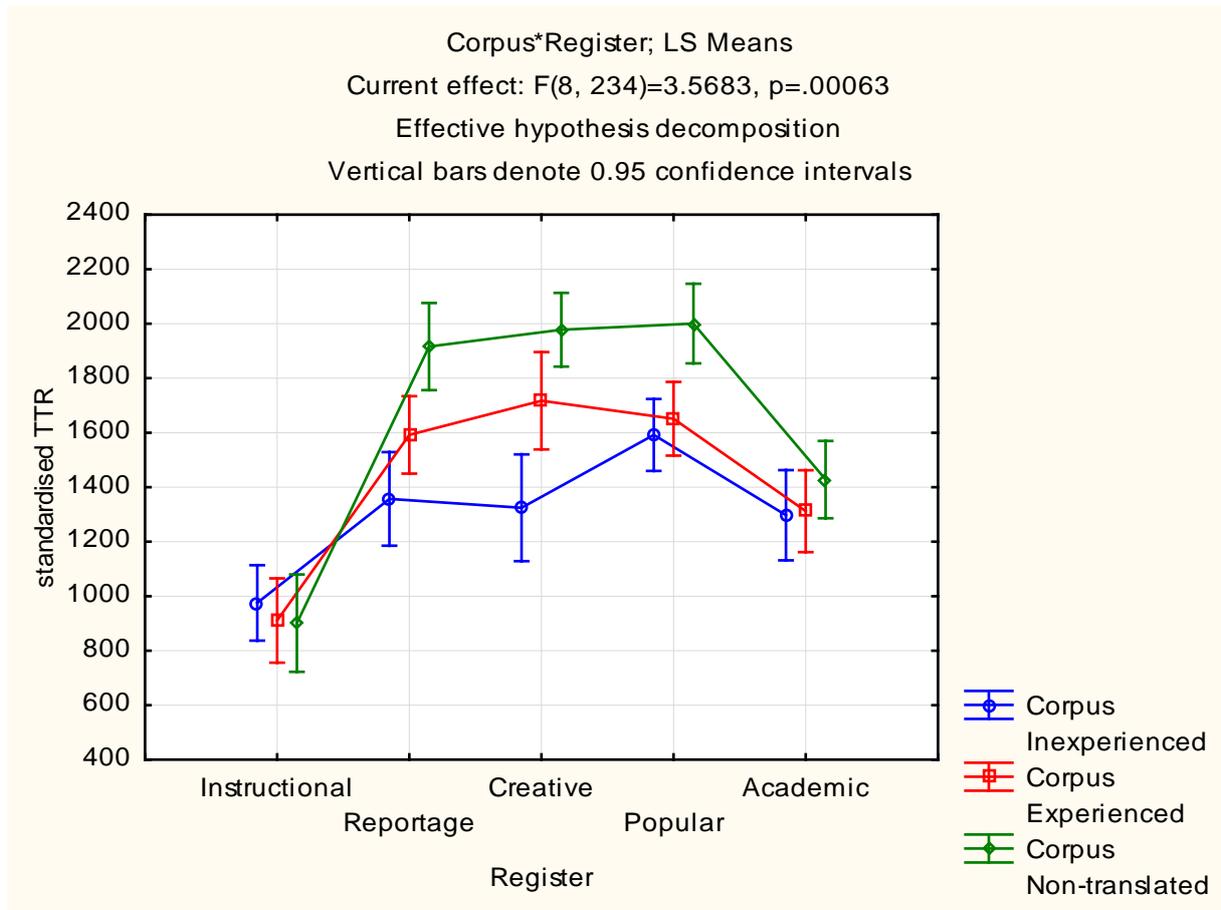
Figure 3: Standardised TTR in the three subcorpora and five registers

9.4   Readability score

The Flesch Reading Ease Test was used to determine a readability score for each text to see if the translations' reading scores level-out across the five registers. 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). This readability index assigns a value on a scale ranging from 0 to 100; the higher the score, the more readable the text is. The standard reading difficulty level is considered to range from 60 to 70 (Williams 2005). Readability scores may be used as an indicator of simplification on the assumption that texts that are easier to read are less complex at the morphological and syntactic levels.

Highly formal texts are often characterised by complex syntax and morphology, which is thought to be due to highly organised discourse (Crystal 1992: 142). This phenomenon is illustrated by the observation of Biber et al. (1999: 24) that academic texts are characterised by morphologically complex vocabulary items, complex noun phrase constructions and frequent passive constructions that contributes to the morpho-syntactic complexity of texts in this register. These characteristics rarely occur in conversation (Biber et al. 1999:24), which is not surprising considering that informal language is loosely structured (Crystal 1992: 142). News reportage has also been found to be characterised by morpho-syntactic complexity whereas fiction takes an intermediate position between reportage and conversation (Biber et al. 1999: 117).

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 of no difference in readability scores across the three subcorpora and five registers cannot be rejected with confidence. Figure 4 shows that there is a clear register effect independent of experience level or translation status. All of the subcorpora have relatively similar values to one another and display a similar pattern for the five registers. The highest scores occur in the creative writing and popular writing registers, the most informal registers in the corpus, and the registers where the highest degree of readability would be expected. The instructional and academic registers have the lowest scores, showing that they are the most challenging to read, which is once again in accordance with what one might expect for formal registers like these. This strong register effect occurs in the translated subcorpora to the same degree as in the subcorpus of non-translations and there is no indication that levelling-out occurred. However, once again some subtler patterns can be noted in that the IT subcorpus consistently shows the highest readability scores across the registers, excluding the reportage register where the score is slightly lower than that of the non-translations.
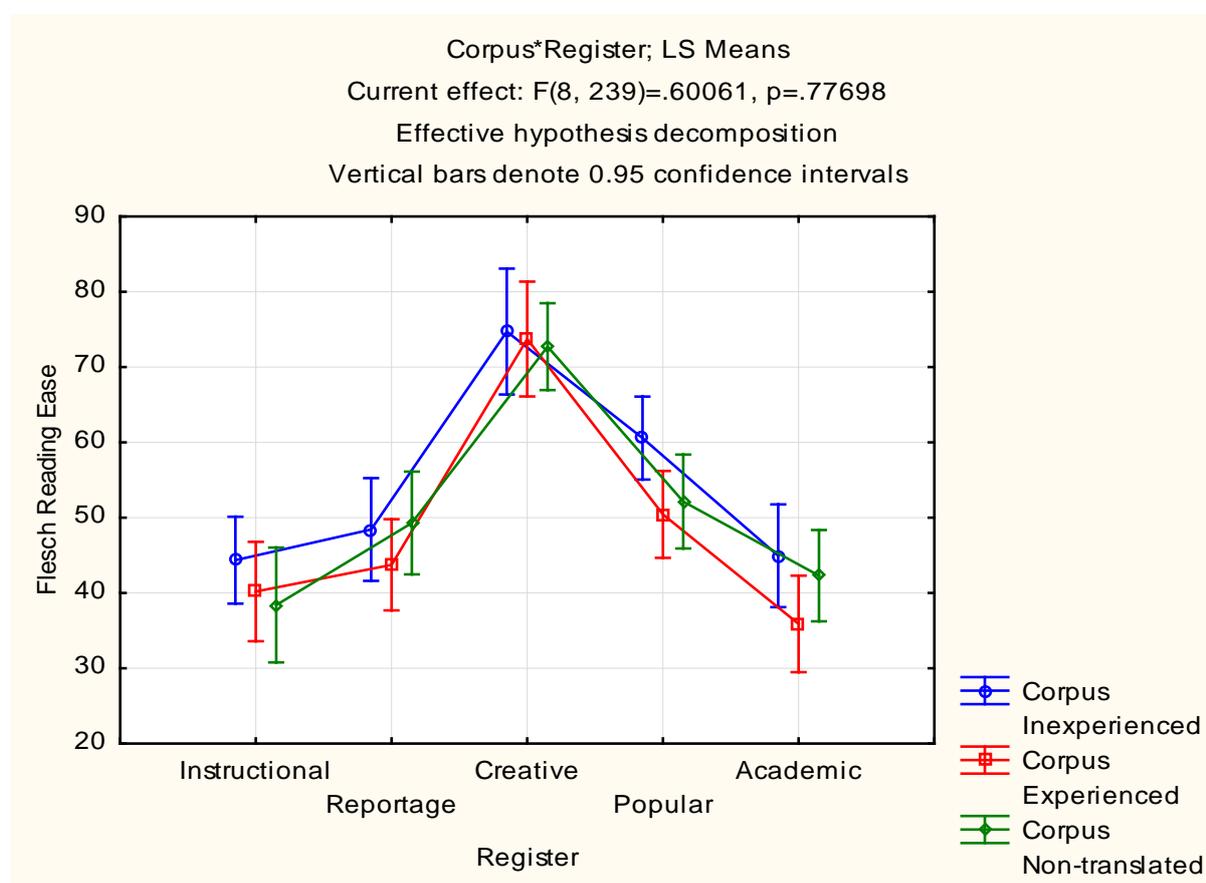
Figure 4: Readability score in the three subcorpora and five registers

9.5   Word length

In terms of register variation, word length is a feature that is strongly associated with register (Kruger 2012: 366). Shorter words that are more general in meaning are more frequent (Biber 1995: 141) than longer words that tend to be more specific, complex and specialised (Biber et al. 1998: 104). According to Biber (1995: 149) longer words are characteristic of high information density as longer words reflect an exact presentation of information content and are more precise

in nature. Longer words, as such, tend to be characteristic of more formal texts with an informational focus, such as technical academic prose (Biber et al. 1998: 149). Shorter word lengths, conversely, are characteristic of informal registers, such as fiction. In translation studies, word length has been used to detect simplification based on the assumption that texts with average shorter word lengths are easier to understand than texts with longer average word lengths, since word length generally correlates with morphological complexity (Kruger 2012: 366).

The results of a factorial ANOVA for the interaction between register and corpus on the measure of mean word length just fails to reach statistical significance (F(8, 239)=1.84, p=0.07), which means that there is some support for differences in register distribution of mean word length across the three subcorpora. From figure 5 it can be noted that the register pattern for mean word length is very similar for the three subcorpora. The more formal academic and instructional registers have the longest word lengths, while the creative writing register has the shortest word length in all three of the subcorpora. However, the NT subcorpus behaves somewhat differently in the reportage and instructional writing registers in comparison to the two translated subcorpora in the sense 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. Even though the two translated subcorpora have a similar pattern over the different registers, it is evident that the IT subcorpus has the shortest word length in general of the three subcorpora. However, there is no evidence that translation causes a levelling-out of word length across registers.
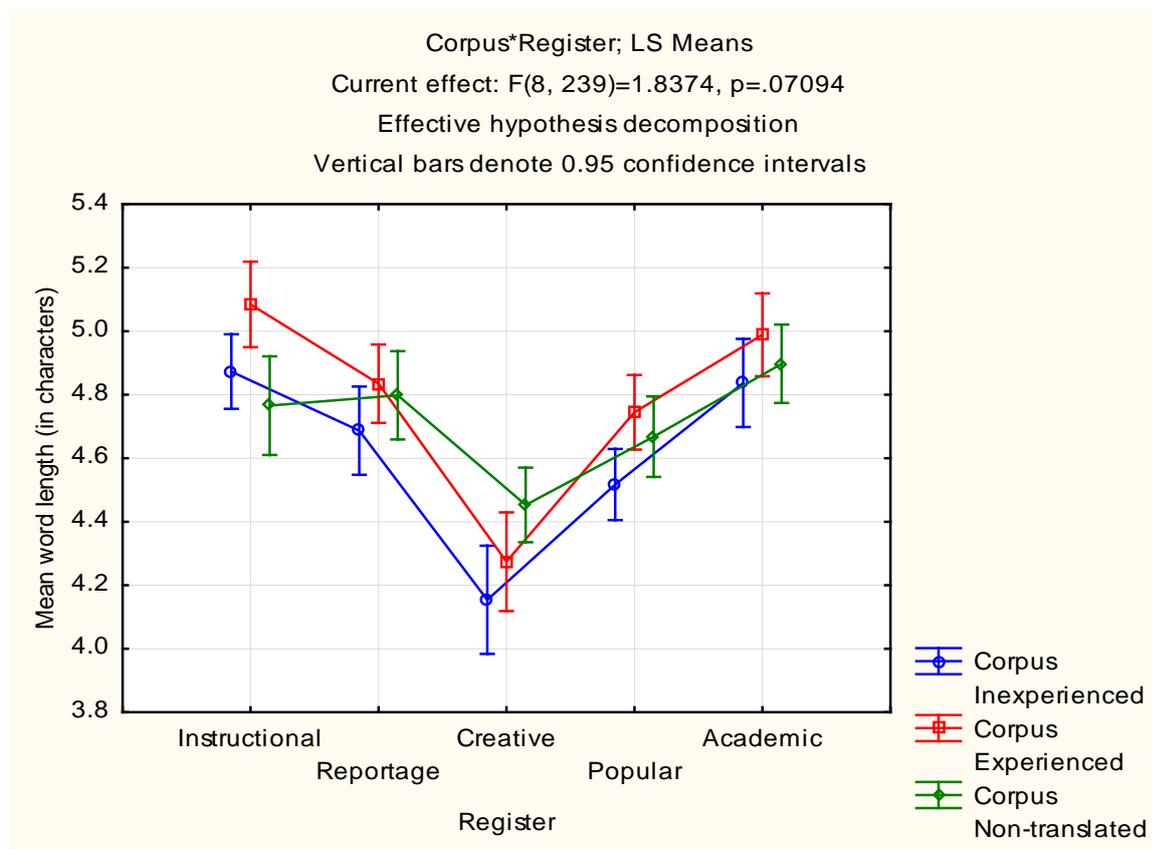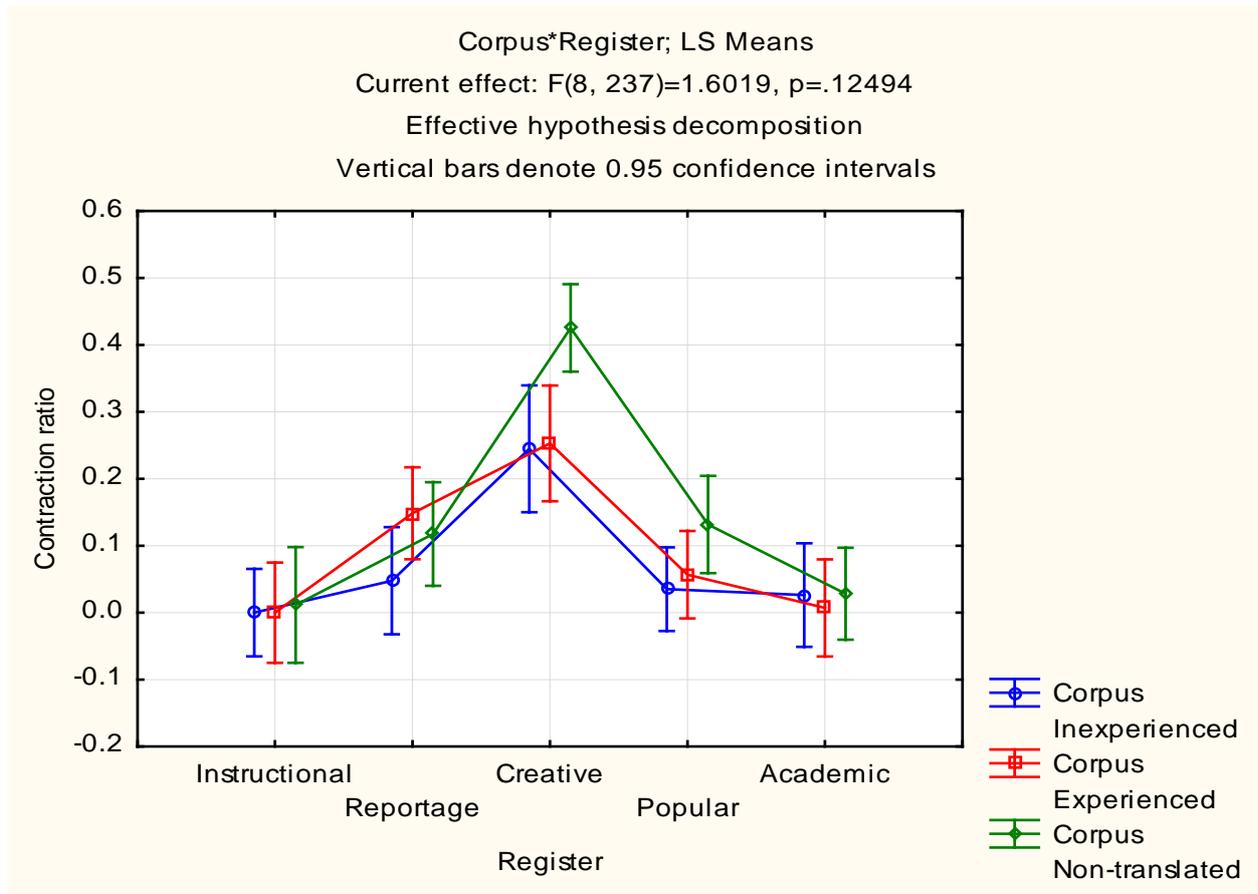

Figure 5: Word length in the three subcorpora and five registers

9.6   Contraction ratio

The use of contracted rather than full forms demonstrates strong register preferences in that contractions are related to register informality (Olohan 2003: 64). Biber et al. (1999: 166), for instance, point out that whereas contracted forms commonly occur in conversations, followed by fiction, the use of the full form is "virtually the only choice in academic prose". Contractions are therefore strongly associated with informal written registers, and are often discouraged in formal written language. In studies of the features of translated language, a lower incidence of contracted forms in translations compared to non-translations is considered to reflect a tendency to conventionalise or normalise to the formal written standard (Kruger and van Rooy 2012: 38).

For this investigation, *not*-negation (contractions with *not*, such as *mustn't* and *don't*) and verb contractions with pronouns (such as *she's* and *they'll*) were combined and used as search terms (see Addendum C). The contraction ratio was used to test whether translators normalise to the written standard by raising the formality of their translations. The results of the factorial ANOVA 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 in the register-specific use of contractions in the three subcorpora cannot be rejected with confidence.

Figure 6 demonstrates that the three subcorpora show similar register preferences for the use of contracted forms. The ET and IT subcorpora demonstrate values very much in the same range that peak in the creative writing register, as is to be expected, where the representation of speech is common. The NT subcorpus has similar values for the academic, instructional and reportage registers when compared with the translation subcorpora but has a visibly higher value for the creative writing register and a somewhat higher value for the popular writing register. While there is a visual suggestion of some degree of levelling-out in the IT subcorpus in terms of the comparably lower frequency of contractions in the more informal creative and popular registers (suggesting inexperienced translators' tendency to apply the conventions of formal writing to informal registers), the statistical analysis does not provide enough support to meet the criterion level of a chance result.

Figure 6: Contraction ratio in the three subcorpora and five registers
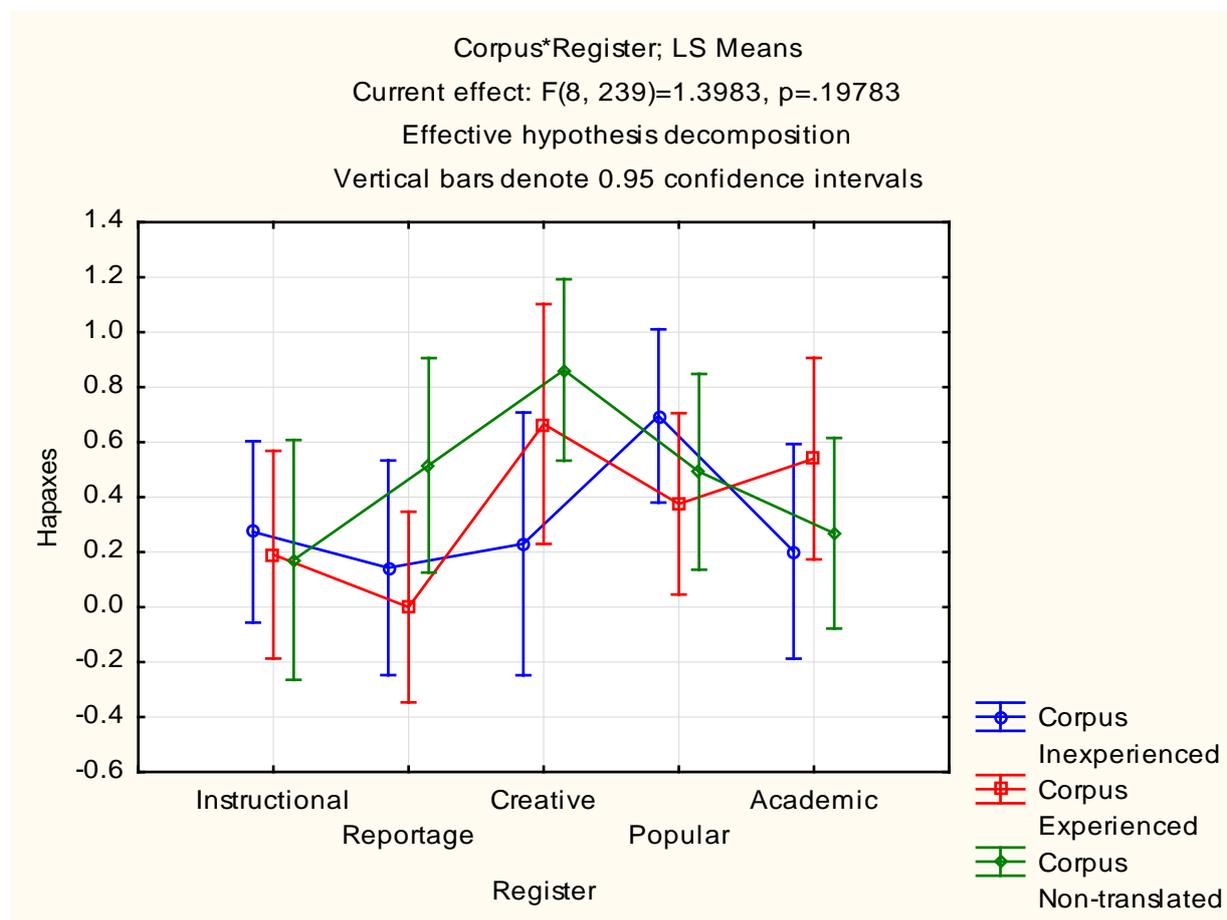
9.7   Neologisms

Neologisms are considered to occur in highly informal registers, typically those that are characterised by a high level of colloquial expression and depart from standard norms (Crystal 1992: 142). Reportage is likely to be caught in a tension between neologisms and more normalised lexical choices. On the one hand reportage, as media whose purpose is to report on news of all types, will likely cover a range of neologisms while its nature as a written genre, on the other hand, may cause it to "under-document developments arising and spreading chiefly in the spoken language" (Mair 2006: 69). For this study, loanwords were included due to the fact that loanwords are considered to be a type of neologism (Algeo 1980: 270).

Neologisms have been used in a variety of different studies as a measurement of normalisation in translation based on the supposition that these lexical-items categories are indicators of lexical creativity. If translation has a linguistically normalising effect, as is proposed by the normalisation hypothesis, the language of translated texts will be more conventional and less creative when compared to non-translations, which will be demonstrated in a lower frequency of non-standard lexical items like neologisms and loanwords in comparison with the frequency of such items in non-translations.

As neologisms are likely to incur infrequently hapax legomena, or word forms that only occur once (Kenny 2011: 143), in the subcorpora were used. The hapaxes were copied into a word-

processing programme and the lexicalised words identified by the spellchecker were deleted. Following this, all abbreviations, spelling errors, acronyms, attested compounds, parts of e-mails and proper nouns were deleted. The words remaining were checked on Oxford Dictionaries Online (www.oxforddictionaries.com) for attested use.

The results of the factorial ANOVA show that there is no interaction effect for the independent variables register and corpus (F(8, 239)=1.40, p=0.20) and the null hypothesis of no difference in the register-related distribution of single-occurrence lexical innovations, therefore, cannot be rejected with confidence. From figure 7 it is evident that there is clearly 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. However, the NT subcorpus follows the prediction that neologisms will be more prevalent in more informal registers than in more formal registers. There is also some visual suggestion of levelling-out in the IT subcorpus, excluding the popular writing register, statistical support is lacking to support the claim.

Figure 7: Neologisms in the three subcorpora and five registers

10.   Conclusion and future work

The present investigation finds little evidence for the first hypothesis that less translation experience reduces register variability in translation that consequently results in levelling-out. There is also little evidence in support of the second hypothesis that translation itself results in a levelling-out of register differences. Therefore, the present investigation finds no evidence for

the claim that translations tend to have a relatively consistent middle register. For all the features studied, register variation was evident in all three subcorpora with neither of the translational subcorpora demonstrating a significantly "flattened out" distribution over the five registers. The findings, in other words, challenge the general assumption that translated texts are more homogenous than non-translations in terms of register variability.

While little evidence was found for the feature of levelling-out, the investigation sheds some light on the relationship between translation expertise and register sensitivity, and also how translations differ from non-translations across different registers for two, namely conjunctive markers and standardised TTR, of the seven features studied. It appears that inexperienced translators struggle in two ways. Firstly, they are inclined to mark the semantic relations between propositions significantly more in translations of informal registers than experienced translators or non-translators. The result, in this regard, suggests that inexperienced translators only explicitate by using conjunctions in registers that are intrinsically more informal in nature. It is possible that they overtly link ideas not only because they misjudge register expectations, but also so as to help them cognitively keep track of the development of the text by using conjunctions to convey logical coherence to themselves.

Secondly, their translations have a smaller vocabulary range than those by experienced translators and non-translators in general. While their smaller vocabulary range is to be expected given their lower language proficiency, the investigation found that their vocabulary range is noticeably more extensive for translations of popular texts than for translations of creative writing. As they have the ability to use a more varied vocabulary for popular texts, the result suggests that inexperienced translators may find creative writing texts more effortful to translate than popular writing texts, which results in them using more repetitive vocabulary when translating the former. In terms of these two features, there is some indication, in other words, that inexperienced translators struggle more with informal registers. A tentative hypothesis that may be postulated is that inexperienced translators (falsely) believe that all instances of written language require formality, as suggested by Pellatt (2012: 158), based on the idea that written texts are more formal in general than spoken registers (Fawcett 1997: 79). In addition, there is also the possibility that translation students transfer the kind of formal writing they are used to using, and for which they are generally rewarded, in the academic context to their translations.

Evidently, experienced translators have a better understanding of the purpose of conjunction markers and how these function across different registers than inexperienced translators. Their register sensitivity, in this regard, is illustrated by the fact that they mark propositions between ideas in different registers in a similar way to non-translators. However, while they have a more extensive vocabulary range than inexperienced translators, their vocabulary is evidently more restricted compared with non-translators. In addition, some of the subtler effects hint towards the register sensitivity of experienced translators in that the features studied followed a similar pattern or distribution in the translations compared with those of non-translations. There is therefore some support for claims made by translation scholars such as Flynn (2013: 18) who argues that translators have the ability to engage directly with different language elements and style typical of a particular genre, and they understand how these various elements play out across languages.

When interpreting the results of the study, it has to be kept in mind that the corpus used for this study is comparatively small. Therefore, it is possible that stronger patterns would manifest in a larger corpus or that even different patterns would emerge is using different features in a larger corpus. In addition, while this study attempted to address the current gap for product-oriented research in investigations of translation expertise, and how translation expertise affects register, it has methodological constraints in terms of the comparable design it used. As pointed out by Bernardini (2011: 12) comparable corpora are characterised by two problematic issues: firstly, they have the disadvantage that they have little explanatory power; and secondly, they depend on the problematic notion of textual comparability. In order to counter the problematic nature of the comparable-corpus methodology, Bernardini (2011: 12) and Saldanha and O'Brien (2013: 69) recommend that the comparable-corpus approach should be combined with a parallel-corpus design that creates bidirectional corpora. This corpus design will enable translation researchers to compare the results generated by their comparable corpora with their parallel text counterparts, which will provide more insight into how translation expertise contributes to register differences. As such, this observation leads to a call for corpus-based studies on translation expertise to incorporate a bidirectional corpus design that will not only produce more reliable results, but will also afford a better understanding of the impact of translation expertise on register.

References

Albir, A.H., F. Alves, B.E. Dimitrova and I. Lacruz. 2015. A retrospective and prospective view of translation research from an empirical, experimental, and cognitive perspective: the TREC network. *The International Journal for Translation and Interpreting Research* 7(1):5–25.

Algeo, J. 1980. Where do all the new words come from? *American Speech* 55(4): 246–277.

Alves, F., A. Pagano, S. Neumann, E. Steiner and S. Hansen-Schirra. 2010. Translation units and grammatical shifts: Towards an integration of product and process-based translation research. In G.M. Shreve and E. Angelone (Eds.) *Translation and cognition*. Amsterdam: Benjamins. pp. 109–141.

Alves, F. and D. Vale. 2011. On drafting and revision in translation: A corpus linguistics oriented analysis of translation process data. *TC3: Computation, Corpora, Cognition* 1(1): 105–122.

Angelelli, C.V. 2009. Using a rubric to assess translation ability: Defining the construct. In C.V. Angelelli and H.E. Jacobson (Eds.) *Testing and assessment in translation and interpreting studies*. Amsterdam: John Benjamins. pp. 13–48.

Angelone, E. 2010. Uncertainty, uncertainty management, and metacognitive problem solving in the translation task. In G.M. Shreve and E. Angelone (Eds.) *Translation and cognition*. Amsterdam: John Benjamins. pp. 17–40.

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

Bernardini, S. 2011. Monolingual comparable corpora and parallel corpora in the search for features of translated language. *SYNAPS – A Journal of Professional communication* 26: 2–13.

Biber, D. 1995. *Dimensions of register variation: A cross-linguistic study.* Cambridge: Cambridge University Press.

Biber, D. 2014. Using multi-dimensional analysis to explore cross-linguistic universals of register variation. *Languages in Contrast* 14(1): 7–34.

Biber, D. and S. Conrad. 2001. Register variation: A corpus approach. In D. Schiffrin, D. Tannen and H.E. Hamilton (Eds.) *The handbook of discourse analysis.* Oxford: Blackwell. pp. 175–196.

Biber, D. and S. Conrad. 2009. *Register, genre and style.* Cambridge: Cambridge University Press.

Biber, D. and E. Finegan. 1994. Register and social dialect variation: An integrated approach. In D. Biber and E. Finegan (Eds.) *Sociolinguistic perspectives on register*. Oxford: Oxford University Press. pp. 315–350.

Biber, D., S. Conrad and R. Reppen. 1998. *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.

Biber, D., S. Johansson, G. Leech, S. Conrad and E. Finegan. 1999. *Longman grammar of spoken and written English*. London: Longman.

Bullon, S., R. Krishnarmurthy, E. Manning and J. Todd. (Eds.) 1990. *Collins Cobuild English Grammar Dictionary*. London: Collins.

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

Chesterman, A. 2011. Translation universals. In Y. Gambier and L. Van Doorslaer (Eds.) *Handbook of translation studies: Volume 2*. Amsterdam: John Benjamins. pp. 175–179.

Colina, S. 2015. *Fundamental of translation*. Cambridge: Cambridge University Press.

Crystal, D. 1992. *An encyclopaedic dictionary of language and languages.* Oxford: Blackwell.

Deeb, Z.A. 2005. A taxonomy of translation problems in translating from English to Arabic. Unpublished PhD dissertation, University of Newcastle. Available online: http://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.420044 (Accessed 14 June 2016).

Ehrensberger-Down, M. and G. Massey. 2014. Cognitive ergonomic issues in professional translation. In W. Schwieter and A. Ferreira (Eds.) *The development of translation competence: theories and methodologies from psycholinguistics and cognitive science.* Cambridge: Cambridge Scholars Publishing. pp. 58–86.

Eskola, S. 2004. Untypical frequencies in translated language: A corpus-based study on a literary corpus of translated and non-translated Finnish. In A. Mauranen and P. Kujamäki (Eds.) *Translation universals: Do they exist?* Amsterdam: John Benjamins. pp. 83–99.

Fawcett, P. 1997. *Translation and language*. London: Routledge.

Flynn, P. 2013. Author and translator. In Y. Gambier and L. van Doorslaer (Eds.) *Handbook of translation studies: Volume 4.* Amsterdam: John Benjamins. pp. 12–19.

Göpferich, S. 2015. Translation competence: Explaining development and stagnation from a dynamic systems perspective. In M. Ehrensberger-Dow, S. Göpferich and S. O'Brien (Eds.) *Interdisciplinary in translation and interpreting process research.* Amsterdam: John Benjamins. pp. 63–78.

Grabowski, L. 2013. Interfacing corpus linguistics and computational stylistics: Translation universals in translational literary Polish. *International Journal of Corpus Linguistics* 18(2): 254–280.

Halliday, M.A.K. and M.I.M. Matthiessen. 2004. *An introduction to functional grammar.* 3rd edition. London: Arnold.

Hansen-Shirra, S. and E. Steiner. 2012. Towards a typology of translation properties. In S. Hansen-Shirra, S. Neumann and E. Steiner (Eds.) *Cross-linguistic corpora for the study of translations: Insights from the language pair English-German*. Berlin: de Gruter. pp. 255–280.

ICE (International Corpus of English). 2012. Available online: http://ice-corpora.net/ice/index.htm (Accessed 22 March 2015).

Kelly, D. 2005. *A handbook for translator trainers: A guide to reflective practice*. Manchester: St. Jerome.

Kruger, H. 2012. A corpus-based study of the mediation effect in translated and edited language. *Target* 24(2): 355–388.

Kruger, H. and B. van Rooy. 2012. Register and the features of translated language. *Across Languages and Cultures* 13(1): 33–65.

Laviosa, S. 1998a. The English comparable corpus: A resource and a methodology. In L. Bowker, M. Cronin and D. Kenny (Eds.) *Unity in diversity: Current trends in translation studies*. Manchester: St Jerome. pp. 101–112.

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

Laviosa, S. 2002. *Corpus-based translation studies: Theory, findings, applications*. New York: Rodopi.

Laviosa, S. 2008. Universals. In M. Baker and G. Saldanha (Eds.) *Routledge encyclopaedia of translation studies*. 2nd edition. London: Routledge. pp. 306–310.

Lesznyák, M. 2008. Studies in the development of translation competence. Unpublished PhD dissertation, University of Pées. Available online: http://nydi.btk.pte.hu/sites/nydi.btk.pte.hu/files/pdf/LesznyakMarta2009_2.pdf (Accessed 14 June 2016).

Mair, C. 2006. *Twentieth century English: History, variation and standardization*. Cambridge: Cambridge University Press.

Martín, R.M. 2014. Situating translation expertise: a review with a sketch of a construct. In W. Schwieter and A. Ferreira (Eds.) *The development of translation competence: theories and methodologies from psycholinguistics and cognitive science.* Cambridge: Cambridge Scholars Publishing. pp. 2–56.

Nesi, H. and H. Basturkmen. 2009. Lexical bundles and discourse signalling in academic lectures. In J. Flowerdew and M. Mahlberg (Eds.) *Lexical cohesion and corpus linguistics*. Amsterdam: John Benjamins. pp. 23–44.

Neumann, S. 2012. Register-induced properties of translation. In S. Hansen-Shirra, S. Neumann and E. Steiner (Eds.) *Cross-linguistic corpora for the study of translations: Insights from the language pair English-German*. Berlin: de Gruter. pp. 191–209.

Neumann, S. 2014. Cross-linguistic register studies: Theoretical and methodological considerations. *Languages in Contrast* 14(1): 35–57.

Nord, C. 2005. *Text analysis in translation: Theory, methodology, and didactic application of a model for translation*. New York: Rodopi.

O'Brien, S. 2015. The borrowers: Researching the cognitive aspects of translation. In M. Ehrensberger-Dow, S. Göpferich and S. O'Brien (Eds.) *Interdisciplinary in translation and interpreting process research.* Amsterdam: John Benjamins. pp. 5–17.

Olohan, M. 2001. Spelling out the optionals in translation: A corpus based study. In P. Rayson, A. Wilson, T. McEnery, A. Hardie and S. Khoja (Eds.) *Proceedings of the corpus linguistics 2001 conference, UCREL Technical papers*. Lancaster: Lancaster University Press. pp. 423–432.

Olohan, M. 2003. How frequent are the contractions? A study of contracted forms in the Translational English Corpus. *Target* 15(1): 59–89.

Olohan, M. 2016. *Scientific and technical translation*. London: Routledge.

Pennsylvania Association of Colleges and Teacher Educators (PACTE). 2009. Results of the validation of the PACTE translation competence model: Acceptability and decision making. *Across Languages and Cultures* 10(2): 207–230.

Pennsylvania Association of Colleges and Teacher Educators (PACTE). 2011a. Results of the validation of the PACTE translation competence model: translation project and dynamic translation index. In S. O'Brien (Ed.) *Cognitive explorations of translation*. London: Continuum. pp. 30–56.

Pennsylvania Association of Colleges and Teacher Educators (PACTE). 2011b. Results of the validation of the PACTE translation competence model: Translation problems and translation competence. In C. Alvstad, A. Hild and E. Tiselius (Eds.) *Methods and strategies of process research: Integrative approaches in translation studies*. Amsterdam: John Benjamins. pp. 317–344.

Pastor, G.C., R. Mitkov, N. Afzal and V. Pekar. 2008. Translation universals: Do they exist? A corpus-based NLP study of convergence and simplification. *Proceedings of AMTA 2008 – 8th Conference of the Association for Machine Translation in the Americas*.

Pellatt, V. 2012. Teaching and learning the importance of ideological awareness for Chinese-speaking trainee translators. In S. Hubscher-Davidson and M. Borodo (Eds.) *Global trends in translator and interpreter training: Mediation and culture*. London: Continuum. pp. 147–160.

Pym, A. 2007. On Shlesinger's proposed equalizing universal for interpreting. In F. Pöchlacke, A.L. Jakobsen and I.M. Mees (Eds.) *Interpreting studies and beyond: A tribute to Miriam Shlesinger*. Copenhagen: Samfundslitteratur Press. pp. 175–190.

Pym, A. 2008. On Toury's laws of how translators translate. In A. Pym, M. Shlesinger and D. Simeoni (Eds.) *Beyond descriptive translation studies: Investigations in homage to Gideon Toury*. Amsterdam: John Benjamins. pp. 311–328.

Redelinghuys, K. & Kruger, H. 2015. Using the features of translated language to investigate translation expertise: A corpus-based study. *International Journal of Corpus Linguistics* 20(3): 293–325.

Saldanha, G. & O'Brien, S. 2013. *Research methodologies in translation studies*. London: Routledge.

Scarpa, F. 2006. Corpus-based quality-assessment of specialist translation: A study using parallel and comparable corpora in English and Italian. In M. Gotti and S. Šaršević (Eds.) *Insights into specialized translation*. Bern: Peter Lang. pp. 155–172.

Schwieter, J. and A. Ferreira. 2014. Underlying processes of L1 and L3 word translation: Exploring the semantic relatedness effect. In W. Schwieter and A. Ferreira (Eds.) *The*

development of translation competence: theories and methodologies from psycholinguistics and cognitive science. Cambridge: Cambridge Scholars Publishing. pp. 87–106.

Scott, M. 2008. WordSmith Tools, version 5. Liverpool: Lexical analysis software. Available online: http://www.lexically.net/wordsmith/version5/index.html (Accessed 15 August 2015).

StatSoft Inc. 2012. STATISTICA (Data Analysis Software System). Version 11.

Steiner, E. 2012. Introduction. In S. Hansen-Shirra, S. Neumann and E. Steiner (Eds.) *Cross-linguistic corpora for the study of translations: Insights from the language pair English-German*. Berlin: de Gruter. pp. 1–17.

Teich, E. 2003. *Cross-linguistic variation in system and text*. Berlin: Walter de Gruyter.

Tirkkonen-Condit, S. 2002. Translationese – a myth or an empirical fact? A study into the linguistic identifiability of translated language. *Target* 14(2): 207–220.

Toury, G. 2012. *Descriptive translation studies and beyond*. Amsterdam: John Benjamins.

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

Van Rooy, B. and I. Esterhuizen. 2011. Die gebruik van konjunksiemerkers by nagraadse studente en gepubliseerde navorsing in die Afrikaanse taal- en letterkunde. *SAALT Journal of Language Teaching* 45(1): 67–86.

Williams, D. 2005. Recurrent features of translation in Canada: A corpus-based study. Unpublished PhD dissertation, University of Ottawa. Available online: http://linguistlist.org/pubs/diss/browse-diss-action.cfm?DissID=8436 (Accessed 14 June 2016).

Xiao, R. 2011. Word clusters and reformulation markers in Chinese and English: Implications for translation universal hypotheses. *Languages in Contrast* 11(2): 145–171.

Xiao, R. and G. Dai. 2014. Lexical and grammatical properties of translational Chinese: Translation universal hypotheses reevaluated from the Chinese perspective. *Corpus Linguistics and Linguistic Theory* 10(1): 11–55.

Yuan, Y. and F. Gao. 2008. Universals of translation: a corpus-based investigation of Chinese translated fiction. In *Proceedings of The International Symposium on using Corpora in Contrastive and Translation Studies.*

Zanettin, F. 2012. *Translation-driven corpora*. Manchester: St. Jerome.

Zanettin, F. 2013. Corpus methods for descriptive translation studies. *Procedia: Social and Behavioural Sciences* 95: 20–32.

Addendum A: Verbs used as search terms for *that*-omission

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

Addendum B: Conjunctive markers

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

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

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			

Addendum C: Contractions and full forms

Verb contraction	Verb full form	*Not*-negation contraction	Full form
could've	could have	aren't	are not
he'll	he will	can't	cannot
here's	here is	couldn't	could not
he's	he is/has	didn't	did not
he'd	he had/would	doesn't	does not
I'll	I shall/will	don't	do not
I'm	I am	hadn't	had not
I've	I have	hasn't	has not
I'd	I would	haven't	have not
it's	it is/was	isn't	is not
it'd	it would	mustn't	must not
it'll	it will	needn't	need not
let's	let us	shan't	shall not
she'll	she will	shouldn't	should not
she's	she is/has	wasn't	was not
she'd	she had/would	weren't	were not
that's	that is/has	won't	will not
they'd	they had/would	wouldn't	would not
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		
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		