A corpus-based study of the mediation effect in translated and edited language

Haidee Kruger
North-West University

This paper presents the results of a study investigating the hypothesis that the recurrent features, or universals, of translated language are primarily the result of a mediation process that is shared among different kinds of mediated language, rather than the particularities of bilingual language processing. The investigation made use of a comparable corpus consisting of a subcorpus of English texts translated from Afrikaans, a subcorpus of comparable edited English texts, and a subcorpus of comparable unedited (and also untranslated) English texts. The frequency and distribution of linguistic features associated with three of the universals of translated language (explicitation, normalisation/conservatism, and simplification) across the three subcorpora were analysed. The study was guided by the hypothesis that the frequency and distribution of linguistic features associated with the universals of translated language would demonstrate similarities in the two subcorpora of mediated text (i.e., the translated and edited subcorpus), as compared to the subcorpus of unmediated text (i.e., the unedited subcorpus). However, the study yields almost no evidence for a mediation effect that is shared by translated and edited language, at least not along the linguistic features investigated. There is, however, evidence for what appears to be a separate translation-specific effect, which seems likely to be more unconscious, more proceduralised and more related to the linguistic level alone. This offers some support for the hypothesis of universals of translated language that are unique to this kind of text mediation specifically. Furthermore, the findings of the study suggest that editing may involve a different kind of mediation effect altogether, which frequently remains invisible in conventional corpus-based studies comparing translated and non-translated language, and which requires further investigation.

Keywords: corpus-based approach, universals of translated language, mediation, edited language, cognitive approaches
1. Introduction

In the mid-1990s, Baker (1993, 1995, 1996) proposed four ‘universals’ of translated language: explicitation, simplification, normalisation/conservatism, and levelling out (see Olohan 2004; Zanettin 2012 for overviews). These ‘universals’ are defined as characteristics that all translated texts share, regardless of the language pair involved, the text type, or the context in which the translation takes place (Chesterman 2004, 3). Baker also suggested a new method for investigating these universals, namely the analysis of comparable corpora of translated and non-translated texts in the same language. In the almost two decades since the original hypotheses, there have been numerous corpus-based studies investigating these proposed universals. While there has been some variation in findings, and the method itself has been questioned (see Pym 2008), findings have indicated support for at least some of the hypotheses (see, amongst others, Baker 2004, 2007; Laviosa 1998; Mauranen 2000; Mutesayire 2004; Olohan and Baker 2000; Pápai 2004).

The question that arises, of course, is why translated language would demonstrate a greater incidence of features related to explicitation, simplification, conservatism and levelling out. As Chesterman (2004, 3) points out, once evidence of shared features of translated language has been found, “we can then look for explanatory hypotheses which would plausibly account for them”. A few such explanatory hypotheses for the universals of translated language have been proposed. Baker (1995) has suggested that such features are the result of translators’ attempt to make texts more accessible to the new readership, while Pym (2005) and Becher (2010) have argued that features associated with explicitation are the consequence of translators’ risk-avoidance strategies. Halverson (2003) has explored the idea that the universals of translated language are the consequence of the specifically bilingual cognitive processing involved in translation.

Ulrych and Murphy (2008) have put forward a case for these universals resulting from the mediation involved in translation, and have argued that these universals should therefore also be visible in other varieties of mediated language (see Bernardini 2007, 14; Chesterman 2004, 10–11 for a similar idea). Mediated language is here understood as instances of language use that have undergone some kind of revision or rewriting after the original production, with a particular audience or purpose in mind (see Ulrych and Murphy 2008, 149–151), typically by a person other than the original text producer, or, less typically, by the text producer herself. Mediated language includes translation, editing, interpreting, subtitling and rewriting. The concept of mediated language may be extended even further, as Chesterman (2004, 10–11) does, into ‘constrained language’ — in other words, language production constrained by various internal and external factors, and which would include, for example, text production by second-language users, or reportage.
In terms of the argument presented in this paper, it is important to note that the concept of the mediation process is understood in primarily cognitive terms. In translation studies, frequently, emphasis is placed on the translator’s role as mediator between languages, audiences, societies and cultures, and analyses of this role have therefore of necessity traditionally focused on texts, languages and cultures (Jääskeläinen 2000, 71), thus drawing the concept of ‘mediation’ into an external frame of reference. This paper does not deny this role, but rather presupposes that this externally defined, norm-based concept of mediation, in whatever form it assumes, is in varying ways internalised by translators and as such forms part of their cognitive processing. In this paper, ‘mediation’ is therefore understood as a cognitive process encompassing various dimensions and aspects. Two of these aspects will be raised in the interpretation of the findings, and therefore require brief explanation here. These two aspects are (a) the conscious/unconscious aspect and (b) the linguistic/social aspect of mediation.

As far as the former is concerned, various proposed models of translation competence or expertise (see, for example, Chesterman 2000; Göpferich 2009, 2011; Muñoz Martin 2009; PACTE 2003, 2005, 2008, 2011; Robinson 2003; Shreve 2006) involve the assumption that some dimensions or components of translation competence are more conscious in nature, in other words, involving primarily declarative knowledge used to solve particular problems in an analytical way, coupled with a high degree of cognitive control and awareness. This kind of cognitive processing is typically comparatively easy to verbalise. Other aspects or components or translation competence are more ‘unconscious’, in the sense that they are highly routinised, automatised and proceduralised, with a lower degree of conscious cognitive control and awareness. This kind of cognitive processing is not as amenable to verbalisation, since it usually operates below the threshold of awareness. In this paper, the distinction between conscious and unconscious processing in translation is extended to the production of mediated text, generally, and some aspects of mediated text-production are therefore foreseen to be more unconscious, routine or proceduralised, and others more conscious, analytical or declarative.

As far as the linguistic/social aspect is concerned, it takes as its point of departure Chesterman’s (2004, 10–11) suggestion that in accounting for the universals of translation, the effects of (at least) two dimensions of cognitive processing should be considered: the cognitive-linguistic as well as the internalised role-perception of translators. Based on this, it is therefore proposed that the social role of language mediators, involving socially sanctioned norms of mediation behaviour (see, for example, Chesterman 1993, 1997; Toury 1995), is internalised and eventually proceduralised by language mediators, and as such impact on the cognitive processing involved in the production of a mediated text. This ‘social’ component
therefore reflects the degree to which language mediators have internalised norms for text production, which may involve various aspects, such as knowledge of language norms and conventions, genre and register sensitivity, an awareness of audience needs, and an awareness of the communication-mediation role of the language mediator. However, there is also a more basic cognitive-linguistic mediation effect, which is the consequence of reprocessing (in the sense of both comprehending and re-producing, in some form) an existing text in a particular linguistic form. The production of mediated text is therefore characterised by cognitive-linguistic constraints — though obviously some kinds of constraints are shared among different kinds of mediated text production, while others are unique. For example, translation obviously involves bilingual processing of a text, while editing is monolingual; but neither of these written forms of mediation typically involves time constraints that may impact text production, as is the case in interpreting or live subtitling. Editing is constrained by working on an existing text, while translation is somewhat ‘freer’ in this respect, since it involves the recreation or reformulation of a text. All these issues will be raised again in the interpretation of findings in Section 4 and 5.

Ulrych and Murphy’s (2008) proposal that translation universals may productively be recast as mediation universals, since translation would share certain mediation-related features with other forms of mediated language, is an intriguing one. However, the findings of their study (Ulrych and Murphy 2008) are somewhat inconclusive. As such, it remains to be seen which effects are shared among and which are unique to different kinds of mediated discourse. Findings in this regard may go some way towards a better understanding of the nature of the cognitive processing involved in translation. The importance of this point is also raised by Chesterman (2004:, 10–11):

The immediate causes of whatever universals there may be must be sought in human cognition — to be precise, in the kind of cognitive processing that produces translations. Constraints on cognitive processing in translation may also be present in other kinds of constrained communication, such as communicating in a non-native language or under special channel restrictions, or any form of communication that involves relaying messages, such as reporting discourse, even journalism. It may be problematic, eventually, to differentiate factors that are pertinent to translation in particular from those that are pertinent to constrained communication in general.

This, then, is the aim of the research project reported on in this paper. The purpose of the project was to investigate the hypothesis that the universals of translated language are the consequence of a cognitive mediation effect that is shared among different kinds of mediated language, and it therefore explores one possible
A corpus-based study of the mediation effect in translated and edited language. The investigation made use of a comparable corpus consisting of a subcorpus of English texts translated from Afrikaans, a subcorpus of comparable edited English texts, and a subcorpus of comparable unedited (and also untranslated) English texts. The frequency and distribution of linguistic features associated with three of the universals of translated language (explicitation, normalisation/conservatism, and simplification) across the three subcorpora were analysed. It was hypothesised that the frequency and distribution of these features would demonstrate similarities in the two subcorpora of mediated text, as compared to the subcorpus of unmediated text.

2. Methodology

This section briefly discusses the composition of the corpus used in the study, the operationalisation of concepts, data collection, and data processing.

2.1 Corpus composition

As pointed out above, the comparable corpus constructed for the purposes of this study consists of three subcorpora: a translation subcorpus, an edited subcorpus, and an unedited subcorpus. All texts included in the corpus are full texts (though in some cases book-length texts were split into chapters), and vary in length from about 1000 words to about 30000 words. Across the corpus, various text types are represented, organised into four registers that were labelled using the standard register labels devised for the International Corpus of English (ICE) (see http://ice-corpora.net/ice/design.htm, 14 August 2012). These four registers are academic, instructional, popular and reportage, comprising the following text types:

- The academic register is represented by extracts from academic books and articles, dissertations and theses.
- The instructional register contains primarily two types of text: administrative writing and policy documents, and educational material (typically school books).
- The popular writing register in all three subcorpora contains some magazine and non-fiction book texts on a variety of topics, intended for popular consumption. This register category also contains popular spiritual texts with a more general informational (rather than persuasive) slant.
- The register of reportage is represented by newspaper articles and news reports for in-house staff publications, as well as some project reports.
The translation subcorpus consists of texts translated from Afrikaans into English in South Africa, in the period 1998 to 2010. The texts were obtained from the language-service offices of two South African universities, the North-West University (Vaal Triangle Campus) and Stellenbosch University, as well as from a private language-service agency. Published texts as well as ephemera are included in the subcorpus, and various levels of translation expertise are represented. However, all translations were produced by either experienced, professional translators generating a substantial part of their income through translation, or by novice translators in their fourth or fifth year of translation training, as part of their internship before leaving the university for a professional position. It was, however, not possible to determine the exact number of translators’ work represented in the subcorpus (it is hoped that this information can be collected at a future date), but the number is estimated to be above 20.

The edited subcorpus consists of texts originally produced in English in South Africa, and edited by professional language editors generating a substantial part of their income through editorial work. As is the case for the translated subcorpus, it was not possible to determine how many different editors produced the texts in the corpus, although it is estimated that the number of editors is similar to the number of translators. The texts included date from 1997 to 2010. The texts were obtained from the same language-service agencies mentioned above, as well as from editors affiliated with the Professional Editors’ Group of South Africa. These texts were received with tracking provided, from which both the original and the edited versions were recovered. (Part of the corpus is therefore also available in parallel form, with unedited and edited versions of the same text.) To supplement registers that were poorly populated in the edited subcorpus (particularly the popular and reportage registers), additional texts were taken from the ICE corpus for South Africa, or ICE-SA (see http://ice-corpora.net/ice/index.htm, 14 August 2012).

The unedited subcorpus is composed of the unedited counterparts of texts in the edited subcorpus. However, a text pair (in other words, the edited and unedited version of the same text) was randomly assigned to either the edited or unedited subcorpus, so a text which is included in the edited subcorpus is never included in the unedited subcorpus, and vice versa.

The final corpus composition is reflected in Table 1. The discrepancies in the sizes of the registers in the subcorpora were unavoidable, given the differences in numbers, length, and types of texts which are typically translated and edited, and which were available to the researcher. However, all statistical calculations were done on standardised scores per 1000 words.
While care was taken to ensure the comparability of the subcorpora, there are some unavoidable differences in the text types included as representative of the four registers in the subcorpora. The possible effects of these differences are taken into account in the interpretation of findings.

2.2 Data collection and processing

2.2.1 Linguistic features used as operationalisations

Table 2 outlines the linguistic features selected as operationalisations of the three universals of translated language investigated in this paper, as well as examples of existing research motivating and utilising the same or similar operationalisations. In the case of operationalisation (f), which, to my knowledge, has not been used before, more detail on the operationalisation is provided in Section 2.2.2.6.

Another factor to be kept in mind regarding these operationalisations (and one which is perhaps not frequently enough raised in existing research), is that the linguistic structures selected as operationalisations are here taken as having the function of explicitation, normalisation or simplification — reflecting the processing involved in translation (or mediation) specifically. However, any given linguistic structure may have multiple functions, as pointed out by Biber et al. (1999, 41–44), including not only functions related to processing and processing constraints, but also discourse-related functions and functions that have to do with social indexing. All these are strongly linked to register. In the context of this paper, with its emphasis on both the linguistic and the social dimensions of mediation, it will therefore be considered important to think of the operationalisations selected not only as indicative of the particular universal of translated language, but also to consider other functions that the linguistic structure may have, particularly as related to register.

2.2.2 Data collection

Data were extracted using WordSmith Tools (Scott 2008). Details about the data collection for each operationalisation are provided below.
Table 2. Features selected for investigation

<table>
<thead>
<tr>
<th>Universal</th>
<th>Subcategory of universal</th>
<th>Linguistic feature used as operationalisation</th>
<th>Examples of previous studies where the particular or a similar operationalisation is motivated</th>
</tr>
</thead>
</table>
| 1. Explicitation | 1.1 More complete/less economical surface realisation in translation | (a) Frequency of use of optional complementiser *that* | Olohan and Baker (2000)  
Williams (2005) |
| | | (b) Frequency of use of full forms versus contracted forms | Olohan (2003) |
| | 1.2 More explicit relations between conceptual propositions in text | (c) Frequency of linking adverbials | Mutesayire (2004) |
| 2. Normalisation or conservatism | (d) Frequency of coinages and loanwords | Bernardini (2011)  
Bernardini and Ferraresi (2011)  
Williams (2005) |
Bernardini (2007, 2011)  
Mauranen (2000)  
Ulrych and Murphy (2008) |
| | (f) Use of inclusive language | – |
| 3. Simplification | (g) Lexical diversity | Laviosa (1998) |
| | (h) Mean word length | Kruger and Van Rooy (2012) |

2.2.2.1 Frequency of use of the optional complementiser *that*. Verbs taking a *that* complement clause were used as search nodes. All the verbs classified by Biber *et al.* (1999, 663–666) as notably common and relatively common verbs were selected for investigation. These verbs are listed in Table 3.

2.2.2.2 Frequency of use of full forms rather than contracted forms. It was decided to investigate verb contractions (with pronouns) and negative contractions which occur at least once in contracted form in the corpus. The list of contractions and full forms investigated is presented in Table 4.
2.2.2.3 Frequency of linking adverbials. A selection of linking adverbials from the six categories defined by Biber et al. (1999, 875–879) was investigated (see Table 5).

2.2.2.4 Frequency of coinages and unlexicalised loanwords. Coinages and loanwords were used as indicators of idiosyncratic and unconventional language use. Since these forms are likely to occur infrequently, it was decided to use hapax legomena as an initial search set. Hapaxes in each subcorpus were extracted. The spelling checker in Microsoft Word was used as a first measure to remove all lexicalised entries. Following this, all proper nouns, acronyms, abbreviations, spelling errors, parts of e-mails, etc. were deleted. The remaining entries were checked using Microsoft Word’s online dictionaries (including Encarta) as well as the Internet. All lexicalised items were removed from the list, and remaining entries were tagged as coinages or loanwords.
### Table 4. Contractions and full forms investigated

<table>
<thead>
<tr>
<th>Contracted form</th>
<th>Full form</th>
<th>Contracted form</th>
<th>Full form</th>
</tr>
</thead>
<tbody>
<tr>
<td>aren’t</td>
<td>are (*) not</td>
<td>shouldn’t</td>
<td>should (*) not</td>
</tr>
<tr>
<td>can’t</td>
<td>can (*) not</td>
<td>that’s</td>
<td>that is</td>
</tr>
<tr>
<td>couldn’t</td>
<td>could (*) not</td>
<td>there’s</td>
<td>there is, there has</td>
</tr>
<tr>
<td>didn’t</td>
<td>did (*) not</td>
<td>they’ll</td>
<td>they will</td>
</tr>
<tr>
<td>doesn’t</td>
<td>does (*) not</td>
<td>wasn’t</td>
<td>was (*) not</td>
</tr>
<tr>
<td>don’t</td>
<td>do (*) not</td>
<td>we’d</td>
<td>we would</td>
</tr>
<tr>
<td>haven’t</td>
<td>have (*) not</td>
<td>we’ll</td>
<td>we will, we shall</td>
</tr>
<tr>
<td>he’s</td>
<td>he is, he has</td>
<td>we’re</td>
<td>we are</td>
</tr>
<tr>
<td>I’d</td>
<td>I had, I would</td>
<td>weren’t</td>
<td>were (*) not</td>
</tr>
<tr>
<td>I’ll</td>
<td>I will, I shall</td>
<td>we’ve</td>
<td>we have</td>
</tr>
<tr>
<td>I’m</td>
<td>I am</td>
<td>who’re</td>
<td>who are</td>
</tr>
<tr>
<td>isn’t</td>
<td>is (*) not</td>
<td>who’s</td>
<td>who is</td>
</tr>
<tr>
<td>it’s</td>
<td>it is</td>
<td>won’t</td>
<td>will (*) not</td>
</tr>
<tr>
<td>I’ve</td>
<td>I have</td>
<td>wouldn’t</td>
<td>would (*) not</td>
</tr>
<tr>
<td>let’s</td>
<td>let us</td>
<td>you’d</td>
<td>you had, you would</td>
</tr>
<tr>
<td>mustn’t</td>
<td>must (*) not</td>
<td>you’ll</td>
<td>you will</td>
</tr>
<tr>
<td>needn’t</td>
<td>need (*) not</td>
<td>you’re</td>
<td>you are</td>
</tr>
<tr>
<td></td>
<td></td>
<td>you’ve</td>
<td>you have</td>
</tr>
</tbody>
</table>

### Table 5. Linking adverbials selected for investigation

<table>
<thead>
<tr>
<th>Enumeration and addition</th>
<th>Summation</th>
<th>Apposition</th>
<th>Result and inference</th>
<th>Contrast and concession</th>
<th>Transition</th>
</tr>
</thead>
<tbody>
<tr>
<td>firstly</td>
<td>in sum</td>
<td>in other words that is</td>
<td>therefore consequently thus</td>
<td>on the other hand</td>
<td>by the by</td>
</tr>
<tr>
<td>secondly</td>
<td>to conclude in conclusion</td>
<td>i.e.</td>
<td>as a result hence as a consequence in consequence</td>
<td>in contrast</td>
<td>incidentally</td>
</tr>
<tr>
<td>thirdly</td>
<td>in summary to summarise(z)e</td>
<td>that is to say which is to say namely</td>
<td>also</td>
<td>alternatively</td>
<td>by the way</td>
</tr>
<tr>
<td>lastly</td>
<td>overall</td>
<td>in other words that is</td>
<td>consequently</td>
<td>anyway</td>
<td></td>
</tr>
<tr>
<td>first of all</td>
<td>all in all</td>
<td>i.e.</td>
<td>therefore consequently</td>
<td>as a result</td>
<td>however</td>
</tr>
<tr>
<td>to begin with</td>
<td>in all</td>
<td>namely</td>
<td>consequently</td>
<td>hence</td>
<td>conversely</td>
</tr>
<tr>
<td>in addition</td>
<td>further</td>
<td>to be exact</td>
<td>as a consequence</td>
<td>as a consequence</td>
<td>instead</td>
</tr>
<tr>
<td>further</td>
<td>furthermore</td>
<td>to be precise</td>
<td>in consequence</td>
<td>in consequence</td>
<td>on the contrary</td>
</tr>
<tr>
<td>likewise</td>
<td>moreover</td>
<td>to be more exact</td>
<td>otherwise</td>
<td>in consequence</td>
<td>by comparison</td>
</tr>
<tr>
<td>similarly</td>
<td>by the by</td>
<td>to be more precise</td>
<td>besides</td>
<td>anyhow</td>
<td>anybody</td>
</tr>
<tr>
<td></td>
<td>incidentally</td>
<td></td>
<td>Nevertheless</td>
<td>never</td>
<td>nevermore</td>
</tr>
<tr>
<td></td>
<td>by the way</td>
<td></td>
<td>still</td>
<td>in any case</td>
<td>in any rate</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>after all</td>
<td>after all</td>
<td>after all</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2.2.2.5 Frequency of common lexical bundles. Biber et al. (1999, 990) define lexical bundles as “recurrent expressions, regardless of their idiomaticity, and regardless of their structural status. That is, lexical bundles are simply sequences of word forms that commonly go together in natural discourse”. For the purposes of this study, lexical bundles were taken as a kind of ‘prefabricated’, conventionalised language unit — and thus indicative of more normalised or conservative language use. In order to determine a set of search terms, it was decided to focus only on trigrams. A list of trigrams in each of the three subcorpora was generated, after which a selection from the full lists was made. The following trigrams were deleted: all trigrams occurring with a frequency of less than 0.01% in each full subcorpus and in fewer than 3% of texts in each subcorpus, as well as all trigrams containing proper nouns and clearly subject-specific words (e.g. ‘university’).

To limit any corpus-specific effects, the reduced lists from all three subcorpora were combined, and trigrams common to all three lists were selected for investigation. This yielded the 25 trigrams set out in Table 6.

Table 6. Trigrams selected for investigation

| a number of | in order to | the development of |
| a result of | in terms of | the end of |
| according to the | it is important | the fact that |
| as a result | it is not | the number of |
| as well as | need to be | the use of |
| be able to | one of the | there is a |
| has to be | part of the | there is no |
| have to be | some of the | well as the |
| | | with regard to |

2.2.2.6 Inclusive language. The use of inclusive language (specifically he/she versus just he, and humankind versus mankind) was investigated as an operationalisation of a conventionalising tendency, on the assumption that both editors and translators as language mediators would be more attuned than original text producers to the convention of avoiding bias and ensuring inclusivity in language. Of course, depending on the viewpoint, the use of non-inclusive language may well reflect a different kind of linguistic and socio-cultural conservatism as well. While issues of inclusivity and gender bias are lingua-culturally constrained, this matter is prominent to a comparable degree in guidelines for English as well as Afrikaans writing (see, for example, Butcher et al. 2006; Miller and Swift 1995; Müller 2003; Schwartz 1995). The confounding effect of lingua-cultural differences in the perception of this matter may therefore largely be excluded for this study, which involves either original English text production, or translation from Afrikaans to English.
2.2.2.7 *Lexical diversity.* Lexical diversity was measured using standardised type-token ratio (TTR) per 1000 words, as computed by WordSmith Tools.

2.2.2.8 *Mean word length.* Mean word length as computed by WordSmith Tools was used, on the assumption that word length can be seen as a measure of morphological complexity. Furthermore, mean word length is also an indicator of lexical specificity. Shorter words are more frequent and more general, while longer words are less frequent and more specific (Westin 2002, 75). Overall, then, word length is an indicator of both complexity and specificity. This is, of course, strongly tied to register, with longer and hence more complex and specific words associated with written, planned, informational and formal registers (see also Biber 1991).

2.2.3 *Statistical analysis*

All values for individual variables, except TTR and mean word length, were standardised to a frequency per 1000 words. Thus, a standardised frequency of each variable was obtained for every individual text. All these values were captured in a spreadsheet and submitted to statistical analysis in Statistica 10 (StatSoft Inc. 2011). In the analysis of results, ANOVA (Analysis of Variance) was used. The analysis in the first instance sought to determine if there were any significant main effects for corpus. Such main effects for corpus would be straightforward evidence for significant differences between the three subcorpora. Where relevant, main effects for register, as well as for the interaction between register and corpus, were also investigated. Furthermore, where the ANOVAs indicated significant differences between the three subcorpora, two-sample t-tests were done to verify between which subcorpora the differences exist. In all cases, two levels of statistical significance were used as guidelines, p<0.05 and p<0.001.

3. Findings and discussion

3.1 Explicitation

3.1.1 *Frequency of the complementiser that*

Using a one-way ANOVA, clauses with complementiser *that* present do not show significant effects for corpus (F(2, 278)=0.04, p=0.96), with frequency in the three subcorpora very similar (see Figure 1).

However, the frequency of clauses where the optional *that* is omitted does show a statistically significant effect for corpus (F(2, 278)=7.46, p<0.05), with the translated subcorpus demonstrating the lowest frequency of omission (see Figure 2), suggesting a greater preference for the more complete surface structure
A corpus-based study of the mediation effect in translated and edited language

However, the edited and translated subcorpora are not similar, with the mean occurrence per 1000 words of the omitted *that*
construction at 0.30 (SD=0.76) in the translated subcorpus, and much higher at 0.90 (SD=1.64) in the edited subcorpus. t-tests confirm this observation, with a statistically significant difference between the translated and edited subcorpus (t(211)=3.57, p<0.001). There is also a statistically significant difference between the edited and unedited subcorpus (t(157)=1.99, p<0.05). However, there is no statistically significant difference between the translated and unedited subcorpus (t(188)=1.41, p=0.16), with the frequency of this feature relatively similar in these two subcorpora.

This result does not confirm the hypothesis that the two subcorpora of mediated text will demonstrate similarities, begging possible explanation for the clear difference between the translated and the edited subcorpus. It seems that there is a translation-specific effect (considering that the editing corpus is also an untranslated corpus), though the effect of source-text constructions on the translations or corpus-design factors may also play a role. An investigation of the role of register may contribute to a more nuanced understanding of the findings.

In the first instance, using a one-way ANOVA, there is a statistically significant effect for register (F(3, 277)=3.14, p<0.05), demonstrating (as expected) clear register-related preferences for the omission of *that*, overall (see Figure 3). Secondly, using a factorial ANOVA there is also a statistically significant effect (F(6, 269)=3.51, p<0.05) for interaction between the categorical variables corpus and register, visible in Figure 4, which indicates that the different subcorpora demonstrate significantly different register-related preferences for *that* omission.

![Figure 3. Mean frequencies of finite complement clauses omitting *that* in the four registers](image-url)
The translated subcorpus, overall, exhibits little variation in preference for the omission of the complementiser *that* according to register, with the reduced form consistently occurring between 0.1 and 0.5 times per 1000 words in the subcorpus. There is some evidence here of a translation-specific cognitive effect that outweighs translators’ more conscious awareness of register-related norms for the omission of *that*. This finding also provides some support for the fourth universal of translated language, levelling out, since register differences are largely neutralised in the translated subcorpus. The unedited subcorpus demonstrates slightly greater register variation for this feature, but it is in the edited subcorpus where register-related differences in the preference for the reduced form is most obvious, with the reduced form much more frequent in the instructional and reportage registers than in any other register in either of the other two subcorpora.

The findings here are difficult to interpret without further qualitative investigation, preferably involving a parallel corpus design. The somewhat different register-related patterns for the omission of *that* evident in Figure 4 may well be the consequence of different cognitive processes involved in the mediated text production of editing and translation. Whereas translation appears to homogenise the frequency of this feature across registers, editing makes register differences somewhat more discernible — perhaps an indication of editors’ particular awareness of register-related norms in this regard. However, this effect is not consistent; in the popular and academic registers the frequency of *that* omission is actually

![Figure 4. Mean frequencies of finite complement clauses omitting *that* in the three subcorpora and four registers](image-url)
quite similar across the three subcorpora. Of course, these findings may also be a consequence of corpus design. One would, for example, have expected greater differences in the frequency of this feature between the academic and popular registers, which are typically at the extremes of the formality scale for written registers. Nevertheless, regardless of the reasons for the differences in register distributions between these features, it is clear that register effects need to be considered when interpreting the frequency of *that* omission as an indicator of explicitation, since register clearly plays a significant role — though the exact nature of this role for different types of mediated language is evidently not straightforward.

### 3.1.2 Frequency of full forms rather than contracted forms

Because the distribution of contracted and full forms was very different in the three subcorpora, the frequency of full forms was calculated as a ratio of the total potential forms that may take a full or contracted form. A one-way ANOVA based on this data yields a significant effect for corpus \( F(2, 277)=4.06, p<0.05 \) (see Figure 5), with the translated corpus preferring the full form around 96% of the time, and preference for the full form higher in both the edited and unedited subcorpus, where this form is used around 98% of the time. The corpus effect is, clearly, between the translated subcorpus and the other two subcorpora, which is confirmed by t-tests. There is a significant difference between the translated and

![Figure 5](image.png)

**Figure 5.** Mean frequency of full forms (expressed as a ratio of all possible instances where a full or contracted form may be used) in the three subcorpora
the edited subcorpus ($t(210)=2.10$, $p<0.05$) and between the translated and unedited subcorpus ($t(187)=2.12$, $p<0.05$), but none between the edited and unedited subcorpus ($t(157)=0.38$, $p=0.71$).

Once again, the clear difference between the translated and edited subcorpora in respect of this feature does not support the hypothesis informing this study. Furthermore, using this operationalisation, the translated subcorpus actually demonstrates a lesser degree of explicitness than the other two subcorpora, which runs counter to the hypothesis of greater explicitness and normalised language use associated with translated language. These findings may well arise from the peculiarities of this corpus (again, more detailed qualitative investigation is necessary). However, it may also be instructive to consider, once again, the role of register, as the use of contracted forms is strongly linked to the formality aspect of register, and this may be an interacting variable.

A one-way ANOVA does, indeed, demonstrate a statistically highly significant effect for register ($F(3, 276)=10.38$, $p<0.001$), with the popular register demonstrating a notably lower preference for the full form than the other three (more formal) registers: around 92% of the time (as compared to the other three registers which prefer the full form between approximately 96% and 100% of the time) (see Figure 6). The (relatively) higher incidence of contractions in the register of reportage is frequently because of the inclusion of direct speech in news reports.

![Figure 6](image-url)

Figure 6. Mean frequency of full forms (expressed as a ratio of all possible instances where a full or contracted form may be used) in the four registers.
However, there is no statistically significant effect for the interaction between corpus and register ($F(6, 268)=1.38, p=0.22$), which means that the register effect is similar across the three subcorpora, suggesting that original text producers, text editors and translators demonstrate similar register-related preferences for the use of full and contracted forms where a choice exists. Taken broadly, the strong effect of register indicates a conscious or a proceduralised awareness of register on the part of language mediators as much as on the part of text producers as far as this linguistic feature is concerned. The corpus and register effects are therefore independent of each other, and the corpus effect (see Figure 5) suggests a clear difference between the translated subcorpus and the other two subcorpora.

### 3.1.3 Frequency of linking adverbials

Using a one-way ANOVA there is no statistically significant effect for corpus for any of the six groups of linking adverbials, nor is there a statistically significant effect for all six categories combined ($F(2, 278)=0.83, p=0.44$), as is evident in Figure 7.

However, even though there is no statistically significant effect, it is evident that the translated subcorpus does make the most use of linking adverbials (mean=2.58, SD=2.41); however, the edited subcorpus makes the least use of linking adverbials (mean=2.19, SD=2.00), and therefore there is little suggestion of similarity between the two subcorpora of mediated language — particularly with

![Figure 7. Mean frequency of all six groups of linking adverbials in the three subcorpora](image-url)
the frequency of this feature in the unedited subcorpus (mean=2.53, SD=2.48) closer to that of the translated subcorpus.

3.2 Normalisation or conservatism

3.2.1 Frequency of coinages and unlexicalised loanwords

As is evident from Figure 8, using a one-way ANOVA there is no statistically significant effect for corpus as far as the frequency of unlexicalised loanwords is concerned (F(2, 278)=0.71, p=0.49). However, loanwords do appear to occur slightly less frequently in the translated subcorpus than in the edited subcorpus (see also Bernardini and Ferraresi 2011), with the unedited subcorpus having the lowest frequency of this feature overall. However, it should be pointed out that, in all three subcorpora, this is a very low-frequency linguistic feature.

The loanwords in the three subcorpora are mostly borrowings from Afrikaans and, to a lesser degree, African languages. Some loanwords from Afrikaans are maatjie (meaning little friend; unedited subcorpus), dubbeltjies (referring to devil’s thorn, a kind of weed; translated subcorpus) and baasskap (meaning to be somebody’s boss or superior; edited subcorpus). Loanwords from the African languages appear to be more limited in all three subcorpora. The names of languages are often used in their loaned form, rather than the Anglicised form, for example isiXhosa and baPedi (unedited and edited subcorpus), and occasionally plant names are borrowings, for example moga, mukwa, motswere and mophane (edited subcorpus). There are also some loanwords from European languages, primarily

![Figure 8. Mean frequency of unlexicalised loanwords in the three subcorpora](image-url)
French (e.g. *flâneur* and *planeur* from the translated subcorpus) and German (e.g. *onkel* from the edited subcorpus), but these are limited to particular texts where they are used either because of terminological requirements in a field of scholarship, or for comic effect in recounting a particular experience.

There is also no significant effect for corpus in the frequency of neologisms (see Figure 9) \( F(2, 278)=0.69, p=0.50 \), with neologisms occurring at a frequency of about 0.05 per 1000 words in the translated and edited subcorpus, and 0.02 in the unedited subcorpus. Neologisms are frequently of a more scientific and terminological nature (e.g. *intrafamily* and *bioresource* from the edited subcorpus and *recurriculising* from the translated subcorpus) although there are some examples where neologisms are more creative and clearly designed for communicative effect, such as in *uncucumber-like* (edited subcorpus) and *The time bomb is Tik-ing* (a word play on *ticking* combined with reference to the drug *tik*, from the translated subcorpus).

The patterns for neologisms and loanwords as linguistic features associated with conservatism therefore show some similarity between the translated and edited subcorpora, with the unedited corpus behaving differently. While the frequency of these features is slightly (almost imperceptibly) lower in the translated subcorpus compared to the edited subcorpus (which is also an untranslated corpus of the kind usually used in comparable corpus designs), the wholly unmediated corpus actually shows a much lower incidence of the idiosyncratic, innovative
forms. It may well be that this is the consequence of corpus design; however, it may also be that the creative dimension of language mediation work plays some role (see Bayer-Hohenwarter 2011; Kenny 2001; Kußmaul 1995), or that there is an effect of interference (in the case of loanwords).

3.2.2 Frequency of lexical bundles
There is a statistically significant effect for corpus \(F(2, 278)=5.61, p<0.05\) in the frequency of trigrams, with the translated subcorpus demonstrating the highest frequency of trigrams (see Figure 10) at a mean of 5.96 per 1000 words (SD=3.08). However, the edited subcorpus is not similar to the translated subcorpus in this regard (mean=4.73, SD=2.29), with a lower frequency than in the unedited subcorpus (mean=5.73, SD=2.29), and the findings therefore do not support the hypothesis of similarities between the two subcorpora of mediated text. This is supported by a two-sample t-test which shows a significant difference in the frequency of this feature between the translated and edited subcorpus \(t(211)=3.20, p<0.05\) and the edited and unedited subcorpus \(t(157)=2.60, p<0.05\). There is no significant difference between the translated and unedited subcorpus \(t(188)=0.51, p=0.61\). The lower frequency of trigrams in the edited subcorpus may possibly be ascribed to editors’ conscious attempt to introduce stylistic variation in texts and to avoid repetition, and is therefore again suggestive of different cognitive processes at work in editing and translation.

![Figure 10. Mean frequency of trigrams in the three subcorpora](image-url)
3.2.3 **Use of inclusive language**

The one-way ANOVA for the straight standardised frequency of forms that may be considered deliberately inclusive and the frequency of the less inclusive, more biased form both demonstrated varied mean frequencies in the three subcorpora, and consequently the frequency of the less inclusive form as a ratio of all instances where a choice between the two forms existed was calculated. Using these ratios, there is no statistically significant effect for corpus \( F(2, 71)=1.74, p=0.18 \), as is evident in Figure 11.

Figure 11 demonstrates that the translated subcorpus is most likely to use inclusive language (with the gender-specific form used only approximately 60% of the time); however, there is no evident similarity with the edited subcorpus, where more than 80% of the time the gender-specific form rather than the inclusive form is preferred. It is possible that editors are more conservative or traditional in preferring the gender-specific form rather than the more inclusive form (which some editors find clumsy or awkward).

3.3 **Simplification**

3.3.1 **Lexical diversity**

Figure 12 shows that standardised TTR for the translated subcorpus is, surprisingly, notably higher at 38.80% unique lexical items per 1000 words, compared to a more similar standardised TTR of 37.26% unique lexical items per 1000 words for

![Graph](image.png)

**Figure 11.** Mean frequency of less inclusive forms (expressed as a ratio of all possible instances where a less inclusive or more inclusive form may be used) in the three subcorpora.
A corpus-based study of the mediation effect in translated and edited language

the edited subcorpus and 36.62% for the unedited subcorpus. The overall effect for corpus is statistically significant (F(2, 277)=4.68, p<0.05).

In this instance, the translated subcorpus seems to be different from the edited and unedited subcorpus. This is supported by a two-sample t-test, which shows a significant difference between the translated subcorpus and the edited subcorpus (t(210)=2.08, p<0.05) and unedited subcorpus (t(187)=2.91, p<0.05), respectively, but not between the edited and unedited subcorpus (t(157)=0.82, p=0.41).

Figure 13 shows a very large and statistically highly significant effect for register (F(3, 276)=54.09, p<0.001), but as can be seen in Figure 14, there is none for the interaction between the variables corpus and register (F(6, 268)=0.87, p=0.52), with standardised TTR following a very similar distribution pattern for the four registers in the three subcorpora.

It therefore appears as though the corpus effect is independent of the register effect, and there is convincing evidence for a difference in the translated subcorpus only — though the difference runs counter to the normal assumption that translated language is more simplified than non-translated language. Possible reasons for this may be found in the particular texts making up the subcorpora. For example, the edited and unedited subcorpora contain some educational texts that frequently contain formulaic, repetitious language, in the form of standardised policy-based outcomes (though this is not reflected in the data on trigrams, where one would expect a correspondence with standardised TTR; see Section 3.2.2). It may also be that the educational texts included in these two subcorpora have lower lexical diversity simply because they are aimed at a younger audience (however, it

![Figure 12. Standardised TTR for the three subcorpora](image-url)
should also be noted that this simplification is not consistently evident in the complexity of word morphology as measured by mean word length; see Section 3.3.2).

Clearly these suggestions are speculative, and an in-depth qualitative analysis of the texts in the corpus is required for a more definitive answer.
3.3.2 Mean word length
Mean word length shows no significant effect for corpus ($F(2, 278)=1.31, p=0.27$), as evident in Figure 15.

However, the data on mean word length is suggestive of similarities between the two subcorpora of mediated text, with the translated and edited subcorpora overall demonstrating marginally shorter word lengths (the very small scale of the graph in Figure 15 should be noted).

4. Summary and interpretation of findings
This section considers the patterns that emerge from the findings, and attempts to find commonalities and differences in the product data that may give some indication of the processes that may have lead to the products, in an effort to formulate explanatory hypotheses on the nature of mediated text production, for further testing.

Overall, the findings do not support the original hypothesis — but they do raise alternative hypotheses. Of the eight features investigated, four demonstrate a statistically significant difference in frequency among the three subcorpora. Two of these are explicitation features (the omission of the complementiser *that*, and the frequency of full versus contracted forms), one a normalisation feature (the frequency of trigrams) and one a simplification feature (lexical diversity as measured by standardised TTR).

![Graph showing mean word length in the three subcorpora](image)

**Figure 15.** Mean word length in the three subcorpora
In interpreting these findings, I would like to compare the information in Table 7 and Table 8. Table 7 shows three broad parameters along which the three subcorpora are similar or different: whether there is some kind of mediation involved in the text production, whether this mediation is monolingual or bilingual, and whether this mediation involves reproducing a new text (either from scratch in the case of original production, or with reference to a source text in the case of translation) or amending an existing text. The latter parameter is therefore concerned with the degree of constrainedness in text production.

Table 8 summarises the pattern of statistically significant differences between the subcorpora, based on the results of the t-tests.

In the first instance, it is evident that there is a consistent difference between the translated and edited subcorpus, with all four features demonstrating a statistically significant difference between the two subcorpora. This means that the main hypothesis informing this study (that the two subcorpora of mediated texts would be similar) is not supported by the findings. In other words, there is no evidence of a significant shared mediation effect between the two mediated subcorpora when compared to each other and also to the unedited subcorpus. However, it should be

Table 7. Three parameters of difference for the three subcorpora

<table>
<thead>
<tr>
<th></th>
<th>Mediation</th>
<th>Within/between languages</th>
<th>Textual constrainedness</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unmediated</td>
<td>Mediated</td>
<td>Monolingual</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bilingual</td>
<td>Producing new text</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Amending existing text</td>
</tr>
<tr>
<td>Translated subcorpus</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Edited subcorpus</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Unedited subcorpus</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

Table 8. Summary of statistically significant differences between subcorpora for four features

<table>
<thead>
<tr>
<th>Feature</th>
<th>Difference between translated and edited</th>
<th>Difference between edited and unedited</th>
<th>Difference between translated and unedited (also untranslated)</th>
</tr>
</thead>
<tbody>
<tr>
<td>That omission</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Contracted versus full forms</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Trigrams</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Standardised TTR</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>
kept in mind that the texts in the edited subcorpus are also untranslated and are, in fact, the kinds of texts that would, conventionally, make up the non-translated component of a comparable corpus design for the investigation of the universals of translated language. Considered from this perspective, the findings for these four features suggest a translation-specific effect. However, this effect is not the consequence of mediation, per se, since the effect is not shared by the two corpora of mediated text. The two subcorpora do differ along two other parameters (see Table 7): the one involves bilingual processing (translation) and the other monolingual processing (editing); and the one is constrained by working directly on an existing text (editing) while the other involves reproducing a new text albeit within the constraining framework of an existing text (translation). The difference between the two subcorpora is therefore clearly the consequence of translation, but this translation effect may be related to either the bilingual processing involved in translation, or the freer production circumstances of translation as opposed to editing — or, of course, a combination of the two factors.

Given that the unedited subcorpus also represents texts that are untranslated, one would expect the same differences also to emerge in a comparison of the translated subcorpus with the unedited subcorpus, if there is a true translation-specific effect. However, here the pattern is less consistent. There is no statistically significant difference between the subcorpora as far as the frequency of the features of that omission and trigrams is concerned. However, there are statistically significant differences between the subcorpora in the use of contracted forms (which occur significantly less frequently in the unedited corpus than in the translated corpus) and in standardised TTR (which is lower in the unedited corpus). While these patterns are difficult to explain, it is clear that, since the same consistent pattern of differences is not observed for the translated versus the two untranslated (i.e., the edited and unedited) subcorpora, the mediating role of editing in the production of (untranslated) texts is something that requires further consideration. The difference, of course, between the edited and unedited subcorpus is that the former is at the extreme end of the constrainedness continuum, whereas the latter (like translation) involves the production of a new text, but is even freer than translation since there is no other existing text guiding its production. It may be possible that this much freer context of production explains why the translation-specific effect is less perceptible in some features when the translated corpus is compared with the unedited corpus.

Clearly the role of editing requires further investigation. The exact nature of this role is difficult to surmise using the current corpus design — for this a parallel design would be more appropriate. There are only two statistically significant differences between the edited and unedited subcorpus: for that omission (which occurs significantly more frequently in the edited than in the unedited subcorpus)
and for trigrams (which occur significantly less frequently in the edited subcorpus). Both these findings actually suggest that in amending texts editors introduce collocational variety, rather than reducing variation in favour of more consistently explicit and standardised language. Mediation in the form of editing therefore does not appear to involve explicitation and simplification to the degree evident in translated language, most likely because editing does not involve the re-processing and reproduction of a text, but only its amendment.

Overall, therefore, the findings of this study do not provide support for the hypothesis that there is a shared mediation effect in translation and editing. Rather, there appears to be an independent, most likely largely unconscious processing effect associated with translation specifically, which is probably the consequence of a combination of the bilingual processing involved in translation, and the active text production involved in translation. In the latter respect, translation is more similar to original text production than editing is (which accounts for the fact that for a number of the features investigated, the patterns are suggestive of similarity between these two subcorpora rather than between the two subcorpora of mediated text).

There is very little evidence of a similarity between the two subcorpora of mediated text, with the only features where these two subcorpora appear to share similarities being the use of loanwords and neologisms (see Section 3.2.1), and mean word length (see Section 3.3.2) — though the findings here are not statistically significant. It may therefore well be that there is a shared mediation effect for translated and edited language, which is simply not perceptible in this study because the features selected for investigation are too infrequent to yield significant results in a relatively small corpus such as this one. It does appear likely, however, that if such a shared mediation effect does exist, it is more related to a conscious awareness of social norms for text production than to a more unconscious linguistic processing effect.

5. Conclusion

The findings of this study raise more questions for investigation than they yield answers. This study has produced almost no evidence for a mediation effect that is shared by translated and edited language. There is, however, more evidence for what appears to be a separate translation-related effect, which seems likely to be more unconscious, more proceduralised and more related to the linguistic level alone. This offers some support for the hypothesis of universals of translated language that are unique to this kind of text mediation specifically.

If one considers the different kinds of cognitive work involved in translation and editing, these findings are not surprising. Editing is bound to involve an engagement with a text that is more limited, more conscious and more aware of
A corpus-based study of the mediation effect in translated and edited language

norm-based expectations for text production in particular contexts, since the text editor is always working on language that has already been produced, rather than (re)producing a text (the exception would be texts that are substantively edited, close to the point of rewriting). Translation, on the other hand, involves active text (re)production, and therefore by its nature implies a larger component of unconscious, proceduralised linguistic production. These differences clearly outweigh the fact that both kinds of text production involve some kind of mediation. Furthermore, the difference between mono- and bilingual text processing must also evidently play a role.

An intriguing further suggestion that emanates from this study is that editing may have its own, peculiar effects. In conventional corpus-based studies of translation universals the non-translated component is almost without exception made up of edited texts. In this study, the fact that what appears to be a translation-specific effect does not replicate exactly for the edited and unedited subcorpora, suggests that editing may affect some changes to unmediated text production that should be considered in studies of translation universals. Investigating the hidden mediating effect of editing on original, unmediated text production may open up productive lines of enquiry for the study of translation universals.

Two other collateral findings emerge from this study. Firstly, this study, like Kruger and Van Rooy (2012), reiterates the importance of considering the register effect in studies of translation universals. Secondly, in a few of the features investigated (e.g. full versus contracted forms, and standardised TTR), the findings run counter to the normal predictions of greater explicitness, normalisation and simplification in translated language. Since this was not the primary focus of this study, possible reasons for these differences were not speculated upon at length. However, clearly such investigation is necessary.

The findings outlined above should be read against the ever present concern with comparability that dogs all comparable-corpus designs. Since comparability is so difficult to ensure, the possibility that the findings are the consequence of flaws in this regard always remains in attendance (Bernardini and Zanettin 2004; Laviosa 1997). Bernardini (2011) has argued for the importance of combining comparable and parallel corpus designs in studies of translated language. In the case of this study, at least two parallel components need to be introduced. The first should consist of the source texts of the texts in the translated subcorpus in order to separate the different kinds of mediation effects more accurately, and to control for the effects of source-text interference (see, for example, Baumgarten et al. 2008; Pápai 2004 for this kind of corpus design), and the second should involve a parallel component in which edited texts are aligned with their unedited counterparts. This will allow for a more nuanced investigation of the actual linguistic and textual effects of editing. Furthermore, additional validation of findings should be
sought by adding a reference corpus (such as the ICE-SA corpus) to the corpus composition, which will, again, contribute to a better understanding of the effects of editing. In fact, an even more complex corpus design may be required, considering that most translated texts are also revised or edited. To investigate the effects of translation without the interference of editing it may in fact be necessary to compare unedited original text production with draft (i.e., unedited and unrevised) translations. Apart from these corpus refinements, the hypotheses formulated above also require experimental testing, so that process-based data may be correlated with product-based findings — one of the ongoing challenges in cognitive translation studies (see Alves et al., 2010).

Notes

1. The term ‘universal(s)’ is, of course, problematic, predominantly as a consequence of the absolutist and universalising tendency innate to it (see Mauaranen and Kujamäki 2004; Toury 2004). While a term like ‘features’ is less totalising and therefore (arguably) preferable, in this paper the term ‘(linguistic) features’ is used to refer to the surface linguistic realisations used as operationalisations to investigate the universals, and the latter term is retained to refer to the broad categories of qualities of translated language, largely for the sake of terminological clarity and ease of usage.

2. It is acknowledged that these models are all very different; however, the aim of this paper is not to formulate a coherent model of translation (or mediation) competence. Rather, reference is made to research on translation competence in order to develop the concept of mediation as it is used in the interpretation of the findings yielded by the corpus-based analysis presented in the paper.

3. The assistance of staff at the university language offices and various editors from the Professional Editors’ Group in making texts available is gratefully acknowledged. In addition, I would like to express my gratitude to my assistants, Karien Redelinghuys and Deidre Duvenage, for preparing the texts for the corpus.

4. The ICE-SA corpus is currently still under construction, and is therefore not yet available in complete form.

5. It is acknowledged that this step allows for the investigation of the number of tokens for each lexical bundle common to the three subcorpora, while obscuring differences in terms of the number of bundle types in the three subcorpora. While such data would be relevant to the hypothesis investigated in this paper, the risk of corpus-inherent differences as a confounding variable was judged too substantial. This method may, however, well prove fruitful for further investigations along these lines, particularly with corpora that are more closely comparable.

6. In principle, interference from the source language is a possible contributing factor, but in practice any effect is unlikely, as Afrikaans has a very limited number of contracted forms, predominantly with the verbs *is* and *het* combined with pronouns (e.g. *hy’s, sy’t*).
References


**Author’s address**

Dr Haidee Kruger  
School of Languages  
North-West University  
Vaal Triangle Campus  
South Africa.

Haidee.Kruger@nwu.ac.za