Translation features in a comparable corpus of Afrikaans newspaper articles

Deirdré Roos

Department of Afrikaans and Dutch, University of Stellenbosch, Private Bag X1, Matieland, 7602, South Africa

E-mail: deirdre.roos@mweb.co.za

1.   Introduction

This article reports some of the findings of a study on Afrikaans translation features in a monolingual comparable corpus of translated and non-translated newspaper articles selected from *Die Burger*[1]. Baker (1993:243) was the first to recommend the use of corpus tools to identify features of translation, which she defined as "features which typically occur in translated text rather than original utterances and which are not the result of interference from specific linguistic systems". Initially the list of translation type features which she thought to be universal consisted of six features, but later she summarised them in four main categories (Ulrych and Bollettieri Bosinelli 1999:235), namely

(i)   explicitation – the tendency to spell out implicit information;
(ii)   simplification – the tendency to simplify the message or language of the target text;
(iii)   normalisation – the tendency to exaggerate patterns and practices typically occurring in the target language; and
(iv)   levelling out – the tendency of translated texts to be "average" and steer away from extremes.

The possibility of these translation features being universal is a much debated topic. Tymoczko (1998:653-6) believes that it is not possible to formulate universals because of changing views, across cultures and through time, of the concept of translation. On this view, it is not feasible to talk about universals if we cannot account for all types of translations and any type of variable (Olohan 2004:20). Others, such as Øverås (1998:3), view the translation features as textual features resorting under Toury's operational norms (cf. Toury 1978:86-88, 1980:53-4, 1995:56-61). There is little consensus as to whether the four above-mentioned translation features are indeed the result of universals or norms (Kenny 2001:53). As corpus translation studies is still in its infancy, no claims can be made yet as to how widespread and how culturally, linguistically and historically independent these features are (Olohan 2004:92).

Baker (1995:235) also suggested a shift away from comparisons between source text and target text to an exploration of how text produced in relative freedom differs from text produced under the normal constraints of translation. Utilising a monolingual comparable

corpus of translated and non-translated language would enable such an exploration. Laviosa-Braithwaite (1995:160-162) suggested a methodology involving certain research hypotheses to investigate these translation features in English. In the study reported here, these hypotheses were used to investigate differences between translated and non-translated language in Afrikaans.

A brief description of the design of the Afrikaans newspaper corpus is given below in Section 2, followed by a discussion of the above-mentioned hypotheses methodology in Section 3. In Section 4 the findings of the investigation into Afrikaans translation features are presented and details are provided on how *WordSmith Tools* (Scott 2008) was utilised in the investigation.

2.   Afrikaans corpus of newspaper articles from *Die Burger*

The Afrikaans comparable corpus of translated and non-translated newspaper articles was created by selecting complete articles from a limited period, namely August, September, and October of 2006. The articles are believed to be representative of ordinary, non-specialised language use, written for the average reader of Afrikaans newspapers.

Laviosa (2002:39-40) argues that comparable newspaper articles are easier to find because translated and non-translated articles can be identified in the same newspaper and therefore one can assume that the readership of the two types of articles is the same. She further argues that dimensions of comparability are larger for newspapers than for other genres because translated and non-translated texts are available in newspapers and each article's subject can easily be identified from the title and subtitle (Laviosa 2002:40).

For the purposes of this study, translated articles were selected first and thereafter non-translated articles from the same domain. I found it easier to first select articles from printed newspapers and then download them from *Die Burger*'s archive on their website (www.dieburger.com). In practical terms, the electronic availability of articles influenced their selection, in the sense that if an article that was selected in the printed newspaper was not available electronically, it was not included in the corpus.

The following domains were included in each of the translated and non-translated subsets: current affairs (3 articles), arts and entertainment (2), business news (12), foreign news (3) and sport, the latter consisting of articles on rugby (5), athletics (2), soccer (6), cricket (8), cycling (1), hockey (2) and golf (1). In total, the corpus consisted of 45 translated and 45 non-translated articles. The total size of the comparable corpus is 36 733 running words (17 450 translated and 19 283 non-translated). According to Biber (in Kennedy 1998:69), a corpus of 2 000 to 5 000 words is large enough to represent a text category such as newspaper articles.

3.   Methodology of hypotheses

Laviosa-Braithwaite (1995:160-162) suggested research hypotheses for explicitation, simplification and normalisation. She views levelling out, Baker's fourth category, as an aspect of normalisation and instead proposes a hypothesis for concretisation, namely that translated texts have a significantly higher frequency of concrete words and/or concrete senses of polysemous words than texts in the source language. To investigate this hypothesis

requires access to the source of the translated texts which is not feasible with a monolingual comparable corpus such as the one used in this study. Table 1 provides Laviosa-Braithwaite's hypotheses as well as the corpus tools she proposed to investigate each aspect.

Table 1. Methodology of hypotheses to investigate the so-called universal translation features

Translation feature	Laviosa-Braithwaite's hypothesis	Tool
Simplification	Translated English texts have a significantly lower type-token ratio (cf. Section 4.1) than original language English texts.	Computer program
	Translated English texts have a significantly higher frequency of superordinates (cf. Section 4.3) than original language English texts.	Word frequency list
	Translated English texts have a significantly lower lexical density (cf. Section 4.2) than original language English texts.	Computer program
Explicitation	Translated English texts have a significantly higher frequency of the optional *that* in reported speech (cf. Section 4.4) than original language English texts.	Keyword in context
	Translated English texts have a significantly higher frequency of repeated, and hence redundant, grammatical items in coordinate structures (cf. Section 4.5) than original language English texts.	Keyword in context
	Translated English texts have a significantly lower frequency of pronouns (cf. Section 4.6) than original language English texts.	Word frequency
Normalisation	Translated English texts have a significantly lower frequency of collocational clashes leading to intentional irony or ad hoc/non-institutional metaphors (cf. Section 4.7) than original language English texts.	Mutual information program and keyword in context

Although the formulation of the hypotheses in Table 1 implies the use of statistical calculations for significance, I used only those statistical calculations provided by *WordSmith Tools* (as explained below). Olohan (2004:86) states that the data themselves do give an indication of a broad tendency and that statistics often do not give more information than the raw frequencies. According to Baker (2004:183), numbers and frequency are only the starting point; they focus our attention on some of the features that could be worthwhile to investigate. For these reasons, statistical significance was not calculated for the Afrikaans corpus.

Laviosa-Braithwaite (1995:160) uses a computer program written by Oliver Jakobs to automatically calculate the type-token ratio as well as the lexical density of texts. Unfortunately, this program is not mentioned by name in her chapter and does not seem to be available commercially. *WordSmith Tools* calculates the type-token ratio, but does not automatically calculate lexical density. More information on the calculation of lexical density in the present study is provided in the following section.

The other tools mentioned in Table 1 are word frequency lists, keyword in context and a mutual information program. *WordSmith Tools* provides word frequency lists, utilises mutual

information, a measure of collocational strength, and indicates the link between two items (McEnery, Xiao and Tono 2006:56), for the keyword in context search in the concordance tool.

The hypotheses methodology predicts that translated texts will differ from non-translated texts in the sense that the former have a lower type-token ratio and a lower lexical density as well as a higher frequency of superordinates, contain the optional *that* in reported speech and have redundant grammatical items in coordinate structures as well as a lower frequency of pronouns and collocational clashes. In the remainder of the article, the findings for the Afrikaans corpus in terms of these aspects are combined with a description of the tools used to investigate them.

4.   Findings for the Afrikaans comparable corpus

As mentioned in the previous section, translated texts are believed to differ from non-translated texts with regard to various aspects that could be investigated by means of a monolingual comparable corpus. The investigation into translation features in the Afrikaans comparable newspaper article corpus was done with the use of *WordSmith Tools 4* (henceforth "WST").

4.1   Type-token ratio

"Types" refer to the number of different words in the text and "tokens" to any sequence of letters with an orthographic space on either side (Baker 1995:236). For example, *die* 'the', *die* 'the' and *'n* 'a(n)' comprise three tokens, but only two types (*die* and *'n).* The type-token ratio (TTR) thus gives an indication of the variety of word forms used in a corpus (Kenny 2001:34) and is a simple indication of the superficial lexical complexity of a text (Munday 1998:4). The TTR is calculated by dividing the number of tokens by the number of types. As TTR is closely related to text length (Chipere, Malvern, Richards and Duran 2004:127), WST uses a standardised TTR similar to the mean segmental TTR of Chipere et al. (2004:128) where the TTR is calculated for continuous segments of the same length. As the translated articles are generally shorter than the non-translated articles, the number of words of the shortest article (179 in the case of this study) can be used as the basis size for a segment for both subsets of the corpus.

WST generates statistics on each subset of the monolingual comparable corpus when word lists for each subset are generated. Table 2 gives the statistics generated by WST with regard to the TTR.

Table 2. Statistics generated by WST for type-token ratio

	Translated articles	**Non-translated articles**
Tokens (running words) in text	17 450	19 283
Tokens used for word list	17 038	18 749
Types (distinct words)	3 711	4 004
Type-token ratio (TTR)	21.78072548	21.35580635
Standardised TTR	60.76723099	60.60123062
Standardised TTR standard deviation	37.37368393	38.40524292
Standardised TTR basis	179	179

Recall that the hypothesis with regard to TTR was that translated (English) texts have a lower TTR than original language (English) texts. The findings for the Afrikaans corpus show that the translated Afrikaans has a slightly higher standardised TTR (60,8) than the original non-translated Afrikaans (60,6). Unlike the hypothesis for English texts then, Afrikaans translated language does not have a lower TTR than non-translated language.

4.2   Lexical density

Lexical density shows the information tempo and information load associated with the use of technical versus general vocabulary, percentage known versus unknown information, the general length of the text and the amount of detail in the description of an event (Baker 1995:237). Lexical density is the percentage content versus function words and is calculated by dividing the number of content words by the total number of words in each subset of the corpus and then multiplying it by 100 in order to calculate a percentage. It is therefore necessary to identify content words, such as nouns, adjectives and verbs, and function words, such as articles and prepositions. WST does not automatically identify content and function words. Parts of speech are not easily identified in the word lists because polysemy and homonymy are not accounted for without context.

De Villiers (1983:55-62) distinguishes between closed and open word classes. Closed word classes are those that are fixed in Afrikaans, namely pronouns, articles, numerals, prepositions, conjunctions and certain adverbs. They form a limited set and do not change often. Open word classes, by contrast, are not fixed and can (and do) change over time – new words are added as nouns, verbs, and adjectives while some words in these classes are no longer used. For the purposes of this study, I classified open word classes as content words and closed word classes as function words to determine lexical density.

An automatic part of speech (POS) tagger was developed for Afrikaans by Pilon (2005) and is available through CTeXT at the North-West University. Although accuracy is not as high as similar taggers for English and Dutch, it is still between 85,87 and 93,69% (Pilon 2005:6, 118). The shortened version of the tagger was used to tag POS into 13 classes, namely nouns, verbs, adjectives, pronouns, adverbs, numerals, articles, prepositions, interjections, conjunctions, residue, unique and punctuation. The tagged data were checked for noticeable mistakes, specifically those involving personal names, homonyms and polysemous words. Tagging is a highly subjective process and the use of a POS tagger would greatly profit replication as well as comparability of the results of future studies on lexical density, specifically for Afrikaans.

Table 3. Statistics for lexical density

	Number of content words	**Total number of words**	**Lexical density**
Non-translated articles	9 299	19 283	48,2%
Translated articles	8 617	17 450	49,4%

The hypothesis with regard to lexical density was that translated texts would have a lower lexical density than original language texts. Unlike the hypothesis for English, I found that the

lexical density of translated texts (49,4%) is slightly higher than that of non-translated texts (48,2%) for the Afrikaans corpus.

4.3   Superordinates

Superordinates are "general words" which include the meanings of various subordinate words, such as *dier* 'animal' (the superordinate) with subordinates or hyponyms *hond* 'dog'*, kat* 'cat'*, leeu* 'lion' and *perd* 'horse' (Carstens 2003:115). Having to work through 3 711 distinct words in the translated subset and 4 004 distinct words in the non-translated subset in search of subordinates, was somewhat overwhelming. I extracted all the nouns from the corpus using the POS-tagged data, and then tried to compare the nouns between the subsets. However, it proved difficult to identify superordinates simply by looking at the word frequency lists or even at the nouns in isolation. Therefore, I selected one comparable article pair, created a word frequency list for both articles and then listed the superordinates and subordinates for the translated and the non-translated article. Admittedly, this led to a highly subjective differentiation, because some subordinates could also be superordinates; for example, *koolstofvesel-onderdele* 'carbon fibre components' could be a subordinate of *koolstofvesel* 'carbon fibre', but could also have subordinates and thus be a superordinate as well. I am therefore careful not to make pronouncements about the hypothesis that translated texts have a higher frequency of superordinates than non-translated texts. Moreover, based on my experience with this aspect of the investigation, I do not believe that corpus tools are suitable for the investigation of the frequency of occurrence of superordinates.

4.4   Optional *that*

In Afrikaans, *dat* 'that' is a neutral conjunction that can be omitted without changing the semantic value of a sentence, especially after a semantically less dominant main clause (Feinauer 1990:116-117). In practical terms, I investigated all the occurrences of *dat* where the sentence could be rewritten without *dat*. For example, (1) can be rewritten without *dat*, as in (2):

(1)   *Pell, die aartsbiskop van Sydney, het gesê <u>hy is dankbaar **dat** daar</u> geen geweld in Australië was in reaksie op die pous se aanhalings uit 'n Middeleeuse teks nie, maar het die gewelddadige protes in ander lande veroordeel.*
        'Pell, the Archbishop of Sydney, said <u>he is grateful **that** there</u> was no violence in Australia in reaction to the pope's quotations from a medieval text, but condemned the violent protests in other countries.'

(2)   *Pell, die aartsbiskop van Sydney, het gesê <u>hy is dankbaar daar was</u> geen geweld in Australië in reaksie op die pous se aanhalings uit 'n Middeleeuse teks nie, maar het die gewelddadige protes in ander lande veroordeel.*
        'Pell, the Archbishop of Sydney, said <u>he is grateful there was</u> no violence in Australia in reaction to the pope's quotations from a medieval text, but condemned the violent protests in other countries.'

This same sentence also involves an example of the omission of optional *dat*, as indicated by "Ø" in (3):

(3)   *Pell, die aartsbiskop van Sydney, <u>het gesê Ø hy is dankbaar</u> dat daar geen geweld in Australië was in reaksie op die pous se aanhalings uit 'n Middeleeuse teks nie, maar het die gewelddadige protes in ander lande veroordeel.*
'Pell, the Archbishop of Sydney, <u>said Ø he is grateful there was</u> no violence in Australia in reaction to the pope's quotations from a medieval text, but condemned the violent protests in other countries.'

The WST concordance tool's keyword in context was used to search the search node *dat* 'that'. The concordance tool's collocation function was then used to identify *verwag* 'expect'*, gesê* 'said' and *verseker* 'reassure/ensure' as verbs used most often with *dat*. Thereafter, *verwag, gesê* and *verseker* were used as search nodes to identify where *dat* was omitted. Table 4 shows the results of these searches.

Table 4. Occurrence of optional *dat*

Verb	Translated	%	Non-translated	%
verwag + dat	3 of 4	75%	4 of 4	100%
Ø	1 of 4	25%	none	0
gesê + dat	4 of 53	7,5%	none	0
Ø	49 of 53	92,5%	75 of 75	100%
verseker + dat	3 of 3	100%	1 of 4	25%
Ø	none	0	3 of 4	75%

As can be seen in Table 4, optional *dat* is sometimes used with *gesê* in the translated subset (7,5% used; 92,5% not used), but *dat* is omitted in all such instances (100%) in the non-translated subset. In the non-translated subset, *dat* is always used when the verb is *verwag* (100%), but this is not the case in the translated subset (75% used; 25% not used). For the verb *verseker* in the translated subset, *dat* is never omitted (100%), but it is omitted in some such instances in the non-translated subset (25% used; 75% left out).

Laviosa-Braithwaite (1995:161) hypothesizes with regard to the optional *that* in reported speech that translated texts have a higher frequency than non-translated texts. I found it to be true for the verbs *gesê* and *verseker* but not for *verwag* in the Afrikaans corpus.

4.5   Redundant grammatical items

To identify redundant grammatical items in Afrikaans, I looked at the use of tautology and pleonasm in the corpora. Carstens (2003:125) sees tautology as the piling up of words of the same meaning, such as *kabeltou* 'cable rope', where *kabel* 'cable' = *tou* 'rope'. He describes pleonasm as a word or idea that is repeated unnecessarily in a certain construction, such as *die skoolhoof van 'n skool* (literally: 'the school head of a school').

I found that the corpus translation studies method could not successfully be applied to study redundant grammatical items in the Afrikaans corpus. One would need to identify search nodes beforehand to be able to extract specific grammatical items from the corpus, and this is not feasible, as tautologies and pleonasms – unlike, for instance, optional *dat* 'that' – do not occur in predictable patterns. In the end, I had to read through every article (which would not be feasible with a larger corpus) and still could not identify redundant grammatical items, possibly because of good subediting to keep articles as short as possible.

I can therefore not make any pronouncement on the hypothesis that translated texts have a higher frequency of repeated, and hence redundant, grammatical items in coordinate structures than non-translated texts. However, I believe that corpus tools can only be applied successfully in the investigation of redundant grammatical items if search nodes are identified beforehand.

4.6   Pronouns

For the investigation into pronouns (words used to indicate things or people without naming them), I used the POS-tagged data for each subset and extracted all the pronouns. Table 5 gives the results. I identified pronouns as a closed word class according to the distinction made by De Villiers (1983:55-62), as mentioned in Section 4.2. Note that one could also use the word frequency list to identify the pronouns and to calculate the frequency of pronouns as a group or as different groups.

Table 5. Frequency of pronouns

	Number of pronouns	**Percentage (of all running words)**
Non-translated articles	1 289	6,7%
Translated articles	1 179	6,8%

The results show that there are more pronouns in the translated subset (6,8%) than in the non-translated subset (6,7%), although there is only a small difference between the two subsets. This could be due to good subediting during which any chance of ambiguity was removed by using explicit nouns. The hypothesis that translated texts have a lower frequency of pronouns than non-translated texts was not borne out in the Afrikaans corpus of newspaper articles.

4.7   Collocational clashes

In WST, collocations are considered to be words occurring near the search nodes, and a distinction is made between two types of collocations, namely "coherence collocates" (words occurring next to each other) and "neighbourhood collocates" (words occurring near one another but not necessarily next to each other). To investigate collocational clashes, I first examined the word frequency lists but noticed that the most frequent words were the function words *die* 'the', *het* 'have', *in* 'in', *van* 'from/of', *en* 'and', *'n* 'a(n)', *is* 'is', *nie* 'not', *te* 'to/too' and *sy* 'she/his'. I then looked at other frequently occurring words present in both subsets of the corpora to find a search node for the concordance search. Frequent words for both subsets were *span* 'team', *mense* 'people/persons', *ander* 'other' and *toets* 'test'. Table 6 shows how often these four search words occur in the translated and non-translated subsets of the corpus.

Table 6. Occurrence of search words in the two subsets

Translated articles: **17 450 running words**	**Non-translated articles:** **19 283 running words**
span (45 times; 0,26% of all words)	*span* (53 times; 0,27% of all words)
mense (27 times; 0,15% of all words)	*mense* (12 times; 0,06% of all words)
ander (29 times; 0,17% of all words)	*ander* (22 times; 0,11% of all words)
toets (36 times; 0,21% of all words)	*toets* (18 times; 0,09% of all words)

As shown in Table 6, instances of *span* 'team' constitute very similar percentages in the two subsets of the corpus. A concordance was created for the search word *span*, and the collocation function was used to ascertain whether collocational clashes could be identified.

There were more combinations with words before *span* (e.g. *die span* 'the team', *sy span* 'his team') in the non-translated subset than in the translated subset, and there were more combinations with words after *span* (e.g. *span se* 'team's', *span in* 'team in/employ') in the translated subset than in the non-translated subset. However, no collocational clashes were noticed. This might be due to the genre under investigation, but I believe the collocation function might prove to be more valuable with larger corpora.

As my results for collocational clashes were inconclusive, I also applied the tagged data in the concordance searches to see if there was any difference between the word classes that *span* combined with in the two subsets. With the tagged data, I could only search coherence collocates (words immediately before or after) but with the keyword in context function access to wider context is provided. I found that *span* was used with parts of speech such as the adjectival particle *bekend* 'known' (**span** *bekend* *gemaak* 'made the team known = announced the team') in the translated subset but not in the non-translated subset. Also, *span* was used more in the translated subset than the non-translated subset before and after pronouns (*sy span* 'his team', *hierdie span* 'this team', *span wat* 'team that/which', *span sy* 'team his'*)* and before prepositions (*span van* 'team of', *span in* 'team in/employ'*)*. Even though this investigation provides interesting data on the differences between translated and non-translated texts, it does not provide evidence for or against the hypothesis on collocational clashes.

5. Conclusion

This article reported on an investigation into translation features in a monolingual comparable corpus of Afrikaans newspaper articles. Following Baker's suggestions regarding categories of translation features and utilisation of monolingual comparable corpora, Laviosa-Braithwaite (1995) developed a methodology consisting of hypotheses that could be used for investigations into translation features in English texts. Although these features are no longer viewed as universal by all, they still provide a basis from which investigations using corpus tools can proceed.

The Afrikaans newspaper corpus was designed with a subset of translated articles and a subset of comparable non-translated articles from *Die Burger,* selected from a limited (3-month) period. The content of the articles covered a wide range of domains, with a total of 36 733 running words in the two subsets combined.

Applying Laviosa-Braithwaite's (1995) suggested hypotheses methodology meant that the differences between the translated and non-translated subset were investigated with regard to TTR, lexical density, as well as the frequency of occurrence of superordinates, the optional *dat* in reported speech, redundant grammatical items, pronouns and collocational clashes.

For this Afrikaans comparable corpus of newspaper articles it was found that Afrikaans translated language differs from non-translated language in terms of a higher TTR, a higher lexical density, a higher frequency of *dat* for the verbs *gesê* and *verseker*, as well as the occurrence of slightly more pronouns. Findings with regard to superordinates, redundant grammatical items and collocational clashes were not conclusive.

The overall conclusion is that, even though most of the suggested hypotheses were shown not to hold true for this newspaper corpus, thus refuting the universal nature of the translation features*,* Laviosa-Braithwaite's (1995) methodology was useful as a starting point for this corpus investigation into Afrikaans translated and non-translated language. Though I found corpus tools difficult to apply to the investigations of superordinates, redundant grammatical items and collocation clashes, the size and genre of this monolingual comparable corpus might have played a role for the latter two aspects. Even though this study could be seen as merely another participant in the "wolf-pack pursuit of translation universals" (Lind 2007:3), I would like to argue that the value of corpus studies such as this one lies in discovering facets of the nature of translated text and translating (in terms of general tendencies) that are not evident to readers of the texts in question.

Note

1.   *Die Burger* was first published in 1915. It is the biggest daily Afrikaans newspaper in the Western Cape with separate editions for the Eastern and Western Cape. It is owned by Media24. Current circulation figures are 80 200 (Monday to Friday) and 99 100 (Saturday) with 511 000 (M-F) and 556 000 (S) readers (mediaclubsouthafrica.com).

References

Baker, M. 1993. Corpus linguistics and translation studies. Implications and applications. In M. Baker, G. Francis and E. Tognini-Bonelli (eds). *Text and Technology. In honour of John Sinclair*. Philadelphia: John Benjamins. pp. 234-250.

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

Baker, M. 2004. A corpus-based view of similarity and difference in translation. *International Journal of Corpus Linguistics* 9(2): 167-193. (Accessed online on 25 January 2008 at: http://researchschool.org/documents/IJCL%20Paper.pdf.)

Carstens, W.A.M. 2003. *Norme vir Afrikaans*. Pretoria: Van Schaik.

Chipere, N., D. Malvern, B. Richards and P. Duran. 2004. *Using a corpus of school children's writing to investigate the development of vocabulary diversity.* (Accessed online on 26 June 2008 at: http://ucrel.lancs.ac.uk/publications/CL2003/CL2001%20conference/papers/chipere.pdf.)

De Villiers, M. 1983. *Afrikaanse grammatika vir volwassenes*. Goodwood: Nasou.

Die Burger. Available online at: www.dieburger.com.

Feinauer, A.E. 1990. Skoon afhanklike sinne in Afrikaanse spreektaal. *South African Journal of Linguistics* 8(3): 116-120.

Kennedy, G. 1998. *An introduction to corpus linguistics*. London: Longman.

Kenny, D. 2001. *Lexis and creativity in translation. A corpus-based study*. Manchester: St. Jerome.

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

Laviosa-Braithwaite, S. 1995. Comparable corpora: towards a corpus linguistic methodology for the empirical study of translation. In M. Thelen and B. Lewandowska-Tomaszczyk (eds). *Translation and Meaning Part 3*. Maastricht: UPM. pp. 153-163.

Lind, S. 2007. Translation Universals (or laws, or tendencies, or …?). *TIC Talk 63* United Bible Societies. (Accessed online on 29 January 2008 at: http://www.ubs-translations.org /tic_talk_bibliographies/past_issues/tic_talk_63_2007/.)

McEnery, T., R. Xiao and Y. Tono. 2006. *Corpus-based language studies: An advanced resource book.* London: Routledge.

Mediaclubsouthafrica.com. 2009. *The press in South Africa.* (Accessed online on 26 November 2009 at: http://www.mediaclubsouthafrica.com/index.php?option=com_contentandview=articleandid=73%3Apressandcatid=36%3Amedia_bganditemid=54.)

Munday, J. 1998. A computer-assisted approach to the analysis of translation shifts. *Meta* 43(4). (Accessed online on 25 January 2008 at: http://www.erudit.org/revue/meta/1998/v43/n4/003680ar.html.)

Olohan, M. 2004. *Introducing corpora in translation studies.* London: Routledge.

Øverås, L. 1998. In search of the third code: an investigation of norms in literary translation. *Meta* 43(4). (Accessed online on 25 January 2008 at: http://www.erudit.org/revue/meta/1998/v43/n4100377sar.html.)

Pilon, S. 2005. Outomatiese Afrikaanse Woordsoortetikettering. Unpublished MA thesis. Potchefstroom: North-West University.

Scott, M. 2008. *WordSmith Tools 4.0.* (Accessed online on 26 June 2008 at: http://www.lexically.net/wordsmith/.)

Toury, G. 1978. The nature and role of norms in literary translation. In J.S. Holmes, J. Lambert and R. van den Broeck (eds). *Literature and translation: New perspectives in literary studies*. Leuven: Acco. (Revised version in Toury 1995: 53-69.)

Toury, G. 1980. *In search of a theory of translation*. Tel Aviv: The Porter Institute for Poetics and Semiotics.

Toury, G. 1995. *Descriptive translation studies and beyond*. Amsterdam: Benjamins.

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

Ulrych, M. and R.M. Bollettieri Bosinelli. 1999. The state of the art in translation studies: an overview. *Textus English studies in Italy* XII(2): 219-242.