

Translating technical texts into Zulu with the aid of multilingual and/or parallel corpora

Rachéle Gauton

Department of African Languages, University of Pretoria, RSA
rgauton@postino.up.ac.za

Gilles-Maurice de Schryver

Department of African Languages and Cultures, Ghent University, Belgium
Department of African Languages, University of Pretoria, RSA
gillesmaurice.deschryver@UGent.be

Abstract

The single biggest problem that translators have to contend with when translating from an international language such as English into a so-called minor language or language of limited diffusion (LLD) such as Zulu, is the lack of terminology in the majority of specialist subject fields. In this article we demonstrate how special-purpose multilingual and parallel corpora can be utilised as a translator's tool in finding suitable term equivalents when translating technical texts into Zulu. Although translators working into languages such as, for example, English and French follow this methodology as a matter of course, it has (to our knowledge) not yet been shown how such a course of action may be applied to the African languages, and specifically to Zulu in this case.

Introduction

The single biggest problem that translators have to contend with when translating from an international language such as English into a so-called minor language or language of limited diffusion (LLD) such as Zulu, is the *lack of terminology* in the majority of specialist subject fields. In languages such as English and French translators are increasingly making use of computerised corpora as a translator's tool to solve this problem of non-equivalence at word level.

The main objective of this article is to demonstrate how a similar methodology could be applied successfully to a previously marginalised language such as Zulu. We will show how special-purpose multilingual and parallel corpora can be

used to find term equivalents when translating technical texts into Zulu. For this purpose, we will present two case studies based on specialised language corpora, one a multilingual corpus dealing with HIV/AIDS and the other a parallel corpus dealing with labour issues.

To begin with, we will discuss the various strategies available to the Zulu technical translator when confronted with terminological gaps in the target language.

Strategies for overcoming the lack of translation equivalents when translating technical texts into Zulu

When confronted with terminological problems, that is, the absence of a particular term in the target language, there are various avenues that the Zulu translator can pursue. Firstly, the existing dictionaries of the language can be consulted, as well as the official *IsiZulu terminology and orthography No. 4* (Department of Education and Training 1993). With the notable exception of a handful of works such as, *inter alia*, Geldenhuys, Viljoen and Mboweni-Marais' *Trilingual business dictionary/Idikshinari yebhizinisi yezilimi ezintathu/Drietalige sakewoordeboek* (1995), Kumalo's *English-IsiZulu linguistic and literary terms/matemu esiNgisi/esiZulu elingwistiksi nelithrisha* (1995) and Rogers' *A basic guide to computer terms/'n Basiese gids vir rekenaarsterme/Incwadi eyisisekelo samagama ekhompinyutha/Hlahlo ya theo ka mareo a khomphuthara* (1999), there is a real dearth of Zulu technical dictionaries, as is indeed the case in all nine of the previously disadvantaged official South African languages (four languages belonging to the Nguni group of languages, namely Zulu, Xhosa, Ndebele and Swati; three languages belonging to the Sotho group of languages, namely Sepedi, Sesotho and Tswana; plus Tsonga and Venda).

Secondly, the translator can source any existing term lists that deal with the particular subject field. Currently, however, the number of existing term lists are woefully inadequate in meeting the technical Zulu translator's needs. (A similar state of affairs also holds in the other official African languages mentioned above.)

Thirdly, the translator can attempt to find parallel texts, that is, any existing Zulu texts in the same subject field as the one in question. (The term *parallel text* is used here as generally understood by translators and is therefore to be distinguished from *parallel corpora* as will be referred to and defined below.) In this regard, the Zulu translator could consult textbooks, pamphlets and other written material, as well as conduct searches on the Internet.

However, in recent years, and particularly in the international languages, this

type of term mining has been done by consulting large electronic corpora, available in machine-readable form. Bowker (2002, 45) states in this regard:

Although corpora are not, in essence, very different from conventional printed parallel texts, their machine-readable form means that they have more to offer translators in a number of respects. For example, when coupled with a corpus analysis tool such as WordSmith, corpora allow translators to consult a wider range of documentation and to easily determine the centrality of patterns found within that range.

If translators have exhausted all of the possibilities listed above and are still unable to find a suitable translation equivalent, they have no other choice than to create a term themselves. When creating or coining new terms, there are a number of guidelines that translators can follow, as explained below.

Firstly, they can attempt to find a translation equivalent for the term in one of the (closely) related African languages, in other words, the Swati translator could consult sources in Zulu, Ndebele or Xhosa. If one of the related African languages should already have a term for the concept in question, the translator is already halfway there, as it is usually much easier translating from one African language into another than translating from a language such as English into an African language. The existence of such a term equivalent in a (closely) related language could stimulate, and in some cases even guide, the translator in deciding on a suitable term creation strategy for his or her target language.

Secondly, when creating terms, the translator has to determine the exact meaning of the term, that is, the concept to which the particular term refers. In the translation of technical texts, this is often a problem as no translator can be expected to be an expert in each and every subject field. For this reason, translators of technical texts in particular tend to specialise in a certain field, for example, information technology, chemistry, mining or aeronautics. In South Africa, however, translators unfortunately do not have the same status as in some other (Western) countries and many translators, especially freelance translators, therefore, translate whichever text types they are called upon to translate. In addition, in South Africa it is often erroneously assumed that anyone who is a primary speaker of a certain language can translate into or out of that language, and that professional training in translation principles and practice is not a necessity. Many translators of technical texts are therefore ill equipped to deal with the particular challenges that this type of translation presents.

Thirdly, the translator has to be completely familiar with the principles of term and word creation in the particular target language – in this case Zulu. When creating terms, the translator should ideally consult specialists or experts in the field that are also primary speakers of the language in question. In this manner,

the translator can standardise the terms to a certain extent as it is clearly of no use if the translator is the only one who can understand and/or will ever use the terms that have been created. If the translator does this type of thorough background work at this stage, the product should be acceptable, as specialist readers will recognise the terminology used.

Lastly, the translator should record all terms (i.e., each source language term and its translation equivalent in the target language) for possible re-use in subsequent translation projects. Particularly important within the South African context is that translators must ensure that they will be in a position to participate in any structures and mechanisms set up to pool, disseminate and ultimately standardise all newly coined terms.

Before proceeding to a discussion of the case studies, the terms *multilingual* and *parallel* corpora need to be defined as used within the context of this study.

Defining *multilingual* and *parallel* corpora

When studying the existing literature on corpus linguistics as compared to corpus-based translation studies in particular, it is clear that there is some disparity amongst researchers in the use of some terms. This inconsistency in the use of terminology, particularly when referring to the concepts *multilingual*, *parallel* and *comparable* corpora, is discussed at length by researchers such as Pearson (1998, 47–48) and McEnery and Xiao (forthcoming) and will not be discussed in any detail here. We follow Baker's (1995) use of the terms *multilingual*, *parallel* and *comparable* corpora, as indicated below.

Multilingual corpora can be defined as a set of two (or more) monolingual corpora in different languages built according to the same design criteria. As Baker (1995, 232) indicates, multilingual corpora allow the researcher to study 'items and linguistic features in their home environment, rather than as they are used in translated text'. Parallel corpora can be defined as original, source language texts in one language and their translated versions in another language. Baker's 'multilingual corpora' are termed *comparable corpora* by Uzar and Walinski (2000) in following Sinclair's (1995, 32) definition, as well as by Zanettin (1998, 2; 2002b, 10) and Bowker and Pearson (2002); and termed *bi-/multilingual comparable corpora* by Johansson (1998, 5) and Laviosa (2003, 101). Baker's *parallel corpora* are referred to as *translation corpora* by, *inter alia*, Johansson (1998, 5) and Zanettin (2002b, 10).

As Bowker (2000) points out, specialised target language corpora can be an extremely valuable resource for translators. Despite the fact that no specialised tools have as yet been developed for extracting translation data from such

corpora, she argues that translators would be remiss not to consult such resources.

As has been demonstrated by Bowker (2000) with respect to French, Uzar and Walinski (2000) regarding Polish, and Taljard and De Schryver (2002) for Northern Sotho (Sepedi), generic corpus analysis tools such as WordSmith Tools (Scott 1999) can be used with great success on various different types of corpora, not only in the extraction of translation-related data, but also for term mining in particular.

As early as 1995, Baker (1995, 223) pointed out that 'corpus-based terminology compilation is now the norm rather than the exception' – a sentiment that rings even more true today.

Methodology

In this article we compare the usefulness of multilingual versus parallel corpora as translation resource and translator's tool. In carrying out this study, raw corpora (i.e., consisting of untagged and unmarked running text) are queried by means of WordSmith Tools, in particular by using the so-called KeyWords function. For contrastive purposes a parallel corpus query tool, namely ParaConc (Barlow 2003), is also used.

The WordSmith Tools key procedure for identifying terminology in each of the various special-field corpora is comparing the frequency of every distinct word-type in those corpora with the respective frequency of the same word-type in general-language or so-called reference corpora – the reference corpora obviously being the larger of the two in each case. Items that display a great disparity in frequency have high (positive) keyness values and are identified as potential terminology, since the disparity implies that those specific items occur with unusual frequency in the special-field corpora.

As reference corpora, the following general-language corpora are used:

- The University of Pretoria Zulu Corpus (PZC), an electronic corpus of five million running Zulu words (tokens) built at the University of Pretoria by Gilles-Maurice de Schryver and Mhlabunzima J. Dlomo, organised chronologically and consisting of a number of sub-corpora stratified according to genre, that is, novels and novelettes; textbooks; short stories, essays and readers; dramas and one-act plays; religious texts; poetry; oral literature, folklore and legends; Internet files and pamphlets (cf. De Schryver and Gauton 2002).
- The University of Pretoria Internet English Corpus (PIEC), an electronic corpus of 12,5 million English tokens culled from the Internet by Rachéle

Gauton, and also consisting of a number of sub-corpora stratified according to genre. Although a relatively large corpus, this can be characterised as a so-called virtual, ephemeral or DIY Web corpus as described by researchers such as Bowker (2002) and Zanettin (2002a; 2002b).

Concerning corpus size, De Schryver and Gauton (2002, 202) point out that when comparing the sizes of *different*-language corpora, one cannot simply contrast the number of words. We counted the number of words in several English texts as well as in their parallel translations into Zulu and found that for every word in Zulu, on average 1.32 words are needed in English. In other words, a corpus of five million Zulu words is roughly equivalent to an English language corpus of 6,6 million words. This difference in the length of orthographic words is caused by the substantial difference between the writing systems of Zulu and English. Zulu has a conjunctive writing system whereby for instance, a whole sentence can be written as one word. For example, a sentence in English with six words 'we have not seen him yet' is written as one word in Zulu, namely *asikamboni*.

We now describe two case studies based on specialised language corpora to determine the feasibility of using such corpora as translation resource and translator's tool in the translation of technical texts into Zulu.

Case study 1: The use of multilingual corpora in the translation of HIV/AIDS terminology

The first case study involves the use of multilingual corpora, consisting of the two corpora as set out below.

Firstly, a Zulu corpus that contains 7 698 tokens, comprising 12 booklets on HIV/AIDS, aimed at children aged between the ages of eight and twelve. Note that this corpus is in essence so-called translated Zulu as the texts have been translated into Zulu from English by a team from the University of Pretoria consisting of Linkie Mohlala and Rachéle Gauton. These texts, furthermore, have a particularly South African focus, dealing with issues such as, *inter alia*, the views of the South African child AIDS activist Nkosi Johnson; mother-to-child transmission of HIV/AIDS; treatment of the disease, and lack of knowledge and myths about HIV/AIDS, peculiar to the South African context; and the use and place of basic health care, clinics and so on in combating the disease.

Secondly, an English corpus of 10 426 tokens. This corpus was culled from the Internet, and contains information on HIV/AIDS aimed at children, and/or frequently asked questions (FAQs) about HIV and AIDS answered in simple terms. When compiling this corpus, we selected only those texts that use plain,

simple language, that is, texts aimed at the non-specialist reader and also included two files specifically dealing with the child AIDS activist Nkosi Johnson. It must be conceded, however, that this corpus is not as tightly focused on the South African context, as is the case with the Zulu corpus.

The next step was to identify terminology in each of these corpora by making use of the KeyWords function of WordSmith Tools, as referred to earlier. In order to eliminate those words that are unusual but infrequent, we selected only those keywords that appear a *minimum* of three times in the multilingual corpora. The two keyword lists were then compared to determine the extent to which such corpora can be used to identify translation equivalents between the two languages. The results of this comparison are indicated below.

For English, 80 per cent of the top 100 keywords are terms, whilst this value drops to 77 per cent for Zulu. The rest are mainly function words, personal names and other non-terms. It must be borne in mind, however, that because of the conjunctive writing system adhered to in Zulu, some of the Zulu keywords actually constitute the same term, preceded by different concordial morphemes. Compare in this regard, for example, the orthographic words in (1) below, all referring to the term AIDS (*ingculazi* in Zulu), but in different contexts:

(1)

N	Orthographic word	South African English
4	<i>nengculazi</i>	(and AIDS)
6	<i>ingculazi</i>	(AIDS)
13	<i>abanengculazi</i>	(who have AIDS)
16	<i>ngengculazi</i>	(about AIDS)
17	<i>yengculazi</i>	(because of/of AIDS)
28	<i>nangengculazi</i>	(and about AIDS)
31	<i>onengculazi</i>	(who has AIDS)
33	<i>lwengculazi</i>	(of AIDS)
38	<i>lengculazi</i>	(of AIDS)
39	<i>yingculazi</i>	(by/it is AIDS)
46	<i>wengculazi</i>	(of AIDS)
67	<i>unengculazi</i>	((s)he has AIDS)
77	<i>ezingculazi</i>	(who have AIDS)
79	<i>benengculazi</i>	(they have/whilst having AIDS)
103	<i>zengculazi</i>	(of AIDS)
114	<i>engculazi</i>	(of AIDS)

Also consider the screenshots in Figures 1 and 2 taken from the corpus query search referred to here.

N	WORD	FREQ	INGCUL-1.TXT %	FREQ	PZOLST %	KEYNESS	P
1	INGCULAZI	287	3.73	0		3,728.2	0.000000
2	HIV	284	3.69	140		3,152.6	0.000000
3	I	300	3.90	8,201	0.16	1,327.1	0.000000
4	NEHNGCULAZI	100	1.30	2		1,277.2	0.000000
5	ISANDULELA	86	1.12	6		1,070.8	0.000000
6	INGCULAZI	66	0.86	60		681.5	0.000000
7	NESANDULELA	46	0.60	1		566.6	0.000000
8	ABANENGCULAZI	45	0.58	0		563.3	0.000000
9	KANYE	129	1.68	7,631	0.15	363.8	0.000000
10	ONESANDULELA	27	0.35	0		349.9	0.000000
11	ABANTU	130	1.69	11,510	0.23	294.2	0.000000
12	IGCIVANE	29	0.38	39		263.2	0.000000
13	ABANENGCULAZI	20	0.26	3		241.4	0.000000
14	MAYELANA	45	0.58	636	0.02	230.5	0.000000
15	NOMA	143	1.86	19,273	0.39	224.2	0.000000
16	NGENGCULAZI	19	0.25	5		221.7	0.000000
17	YENGCULAZI	16	0.21	0		207.3	0.000000
18	USAL	14	0.18	0		161.4	0.000000
19	NGESANDULELA	14	0.18	0		161.4	0.000000
20	UNESANDULELA	13	0.17	0		168.5	0.000000
21	JOHNSON	13	0.17	9		138.7	0.000000
22	NOHLE	12	0.16	4		137.5	0.000000
23	AFRIKA	31	0.40	832	0.02	137.2	0.000000
24	OMAZIMBA	15	0.19	35		133.4	0.000000
25	KWAZOCANSI	10	0.13	1		122.9	0.000000
26	LOKWELAPHA	12	0.16	26		108.2	0.000000

FIGURE 1: Screenshot of the results of the corpus query carried out on the Zulu HIV/AIDS corpus by means of the WordSmith Tools KeyWords function

N	WORD	FREQ	CUL-1.TXT %	FREQ	PZOLST %	KEYNESS	P
1	NENGCULAZI	100	1.30	2		1,277.2	0.000000
2	INGCULAZI	66	0.86	60		681.5	0.000000
3	ABANENGCULAZI	20	0.26	3		241.4	0.000000
4	NGENGCULAZI	19	0.25	5		221.7	0.000000
5	YENGCULAZI	16	0.21	0		207.3	0.000000
6	NANENGCULAZI	8	0.10	0		103.7	0.000000
7	ONENGCULAZI	8	0.10	2		93.7	0.000000
8	LWENGCULAZI	7	0.09	0		90.7	0.000000
9	LENGCULAZI	7	0.09	2		81.2	0.000000
10	YINGCULAZI	8	0.10	9		80.2	0.000000
11	WENGCULAZI	6	0.08	0		77.7	0.000000
12	UNENGCULAZI	5	0.06	2		56.4	0.000000
13	SINENGCULAZI	4	0.05	0		51.8	0.000000
14	BENENGCULAZI	4	0.05	0		51.8	0.000000
15	ZENGCULAZI	4	0.05	5		39.5	0.000000
16	ENGCULAZI	3	0.04	0		38.9	0.000000

FIGURE 2: Instances of *ingculazi* (AIDS) proposed by the WordSmith Tools KeyWords function

When cross-comparing the first 100 keywords for the two multilingual corpora, with both lists calculated by the WordSmith Tools KeyWords function, we found that for 71 per cent (i.e., 55/77) of the Zulu terms, translation equivalents could be found within the first 100 English key words. Furthermore, 10 terms were found, not in the first 100 English keywords, but within positions 100–153. Only 12 Zulu terms, predominantly all with a particularly South African focus, had no counterparts within the first 153 English keywords. These terms focus on issues such as:

- the role of primary health care and health care workers in the treatment/prevention of AIDS, for example, *ikliniki* 'clinic', *odokotela nonesi* 'doctors and nurses';
- AIDS awareness campaigns, for example, *iribhoni* '(AIDS) ribbon commonly worn and displayed in South Africa', *gwema* 'avoid';
- the dispelling of AIDS myths peculiar to the South African context regarding the spread of and possible cures for the disease, for example, *iqiniso/amaqiniso* 'truth(s)', *ososayensi* 'scientists';
- living with HIV/AIDS, for example, *ukwesekela* 'support', *amalungelo* 'rights', *indawo ephephile* 'place of safety'.

Recall that, unlike the Zulu special-purpose corpus, the English corpus does not focus specifically on the (South) African AIDS situation. It is therefore highly likely that should this be remedied, there could be an even better 'fit' between the corpora as far as translation equivalents are concerned.

However, when using these corpora to find Zulu translation equivalents for English terms, a different situation holds. For only 36 per cent (i.e., 29/80) of the English terms occurring within the first 100 key words, we found translation equivalents within the first 100 Zulu keywords. With a few exceptions, the rest of the translation equivalents occur within positions 100–205 in the Zulu keyword list. The reason for this is again clearly due to the conjunctive writing system of Zulu, where there is more than one equivalent for one and the same English term, depending on the context in which it is used. See again example (1) in this regard. This would seem to indicate that, when using multilingual corpora to find Zulu translation equivalents for English terms, the 'fit' between the corpora regarding the position occupied by a particular term is not that snug, with the result that the translator needs to look a bit further afield, that is, further down in the Zulu keyword list, to find the necessary translation equivalent.

Case study 2: The use of parallel corpora in the translation of labour terminology

In the second case study, we make use of parallel corpora dealing with labour issues to determine the usefulness of these types of corpora as resource for the translation of technical texts into Zulu in this subject field.

The two corpora used are an English text of 1 479 tokens and its Zulu translation containing 1 004 running words. These small corpora illustrate the type of micro-scale corpus work that practising translators engage with on a daily basis. They are typical of mini, disposable, virtual or DIY corpora as discussed by researchers such as Bowker (2002), Zanettin (2002a; 2002b) and Maia (2003). Zanettin (2002b, 12) states in this regard that ‘the less “robust” (i.e., the more “virtual”) corpora are the most truly professional type, with reference to translators, since they are “rough-and-ready” products created for a specific translation project’.

As can be expected, the ‘fit’ between these parallel corpora is better than in the case of the multilingual corpora discussed earlier. When doing a KeyWords search on these small corpora, 24 Zulu and 37 English terms result. On comparing these terms, we found that 92 per cent (i.e., 22/24) of the Zulu terms have translation equivalents in the English list, whereas only 43 per cent (i.e., 16/37) of the English terms have translation equivalents in the Zulu list.

Querying these corpora by making use of a parallel concordancing tool such as ParaConc results in an increase in the success rate with which translation equivalents are identified. This is to be expected in a program such as this as the sentences of the two texts can be aligned, which is not the case when doing a KeyWords search using WordSmith Tools.

By applying some of the main functions of a tool such as ParaConc (e.g. the various parallel search functions, the so-called hot word search function that makes use of frequency information to suggest possible translation equivalents for the searchword) to a set of (aligned) parallel corpora, there will be a 1:1 correlation between the source term and its possible translation equivalent(s) (cf. Barlow 2002). However, when using aligned parallel corpora as translation resource, it would serve the translator better to rather make use of a translation memory tool that is purpose-built for the task.

Nevertheless, ParaConc’s corpus frequency function can also be used by the translator as a quick way to obtain a better result in the identification of translation equivalents than that obtained when querying a parallel corpus by means of a tool such as WordSmith Tools. The added advantage is that when utilising ParaConc’s corpus frequency function the parallel corpora do not

necessarily have to be aligned, which means that no more preparation time is needed than when applying, for example, WordSmith-Tools to such a corpus.

By means of ParaConc's corpus frequency function, we attempted to extract the first 100 words that appear a *minimum* of three times in the two parallel corpora dealing with labour issues. This query returned a total of 98 English and 57 Zulu words. Within this list, we identified 56 English and 41 Zulu terms, which constitute 57 per cent (56/98) of all the English words and 72 per cent (41/57) of all the Zulu words in the two lists. Consider in this regard also the screenshot from the ParaConc search in Figure 3.

English (United Kingdom)			English (South Africa)		
Count	Fct	Word	Count	Fct	Word
78	5.3942%	the	40	4.0000%	ama/hosikazi
77	5.3250%	and	29	2.9000%	esamwu
61	4.2185%	women	15	1.5000%	kufanele
57	3.9419%	of	9	0.9000%	akuna
50	3.4578%	to	9	0.9000%	emisebenzini
36	2.4896%	samwu	9	0.9000%	ukuthi
29	2.0055%	in	8	0.8000%	inhlango
22	1.5214%	all	8	0.8000%	roma
21	1.4523%	be	8	0.8000%	onke
20	1.3831%	must	7	0.7000%	kuthi
20	1.3831%	that	7	0.7000%	yabasebenzi
14	0.9682%	for	5	0.5000%	kanye
13	0.8990%	a	5	0.5000%	lolu
11	0.7607%	as	5	0.5000%	sithi
11	0.7607%	at	4	0.4000%	amakungelo
11	0.7607%	demand	4	0.4000%	bonke
11	0.7607%	is	4	0.4000%	emisebenzini
10	0.6916%	on	4	0.4000%	ezaphathelene
10	0.6916%	shall	4	0.4000%	indlele
10	0.6916%	this	4	0.4000%	isamwu
10	0.6916%	with	4	0.4000%	izindlele
10	0.6916%	workplace	4	0.4000%	kwamakhosikazi
9	0.6224%	by	4	0.4000%	ngobulili
9	0.6224%	work	4	0.4000%	njalo
8	0.5533%	aids	4	0.4000%	phanibili

FIGURE 3: Screenshot of the results of the corpus query carried out on the parallel labour corpus by means of ParaConc's corpus frequency function

When studying these frequency lists with a view to identifying translation equivalents, we found that, whilst 100 per cent of the Zulu terms had translation equivalents in the English list, only 52 per cent (29/56) of the English terms had translation equivalents in the Zulu list.

Again this would seem to confirm what was found in the first case study, namely that when using corpora (whether multilingual or parallel) to find Zulu

equivalents for English terms, the translator needs to access a longer list of Zulu terms than is the case when translating from Zulu into English.

As Bowker (2000, 24) rightly observes:

To successfully use a generic corpus analysis tool for extracting translation-related information [...] translators must bring all their conceptual, linguistic and textual knowledge to bear on the task at hand. Much of the initial work may be based on trial and error, and strategies that work for a particular language pair or a specific text, text type, or subject field may not be applicable in other situations.

Conclusion

We have indicated here how special-purpose corpora (whether multilingual or parallel) can be used with a certain degree of success as a resource and tool for translators of technical texts into Zulu. As far as we are aware, this type of application has not been documented (and/or has conceivably not been used) for Zulu before. This is presumably largely because researchers as well as professional translators have not had access until now to a large general corpus of Zulu that could be used as a reference corpus when determining keyness values (cf. the WordSmith Tools tests), and/or because parallel concordancing software packages (cf. the ParaConc tests) have not been in widespread use in South Africa. As both these resources, namely corpora and corpus query software, are now available to Zulu translators, we believe that they should be used fully, whilst taking into account the allowances that need to be made for Zulu's conjunctivism when attempting to find Zulu translation equivalents for English terms.

With regard to the methodologies themselves, we have seen that the best results are obtained when the translator has access to parallel corpora in a particular subject field, at which point these corpora can simply be fed into parallel concordancing software and with or even without sentence alignment, translation equivalents may then rather easily be identified. The corpus functions as a kind of translation memory in such cases. When the translator makes use of a set of multilingual special-purpose corpora as a translator's tool, then he needs a corresponding set of general-language reference corpora in addition, and any software application that can calculate keyness values.

With reference to the Western context, Baker (1995, 224) paraphrases Sager as follows: 'Terminology compilation is now firmly corpus-based. The desire to construct abstract and neat conceptual systems has given way here to the practical need of addressing what happens in real life. Terms are therefore no longer extracted from previous lists but are rather drawn from a representative corpus of authentic texts held in electronic form.'

With local text corpora and language-independent corpus technology now available to South African translators, it is not only possible but indeed eminently feasible for those translating even highly technical texts into a previously disadvantaged and marginalised language such as Zulu, to make use of computerised multilingual and parallel corpora as translation resource and translator's tool.

References

- Baker, M. 1995. Corpora in translation studies: An overview and some suggestions for future research. *Target* 7 (2):223–243.
- Barlow, M. 2002. ParaConc: Concordance software for multilingual parallel corpora. In *Language resources for translation work and research, LREC 2002 Workshop Proceedings, Las Palmas de Gran Canaria*, ed. E. Yuste-Rodrigo, 20–28, Available from <http://www.ifi.unizh.ch/cl/yuste/postworkshop/postworkshop.htm>.
- . 2003. *ParaConc: A concordancer for parallel texts*. Houston, TX: Athelstan. See for this software also <http://www.athel.com>.
- Bowker, L. 2000. Towards a methodology for exploiting specialized target language corpora as translation resources. *International Journal of Corpus Linguistics* 5 (1):17–52.
- Bowker, L. 2002. Working together: A collaborative approach to DIY corpora. In *Language resources for translation work and research, LREC 2002 Workshop Proceedings, Las Palmas de Gran Canaria*, ed. E. Yuste-Rodrigo, 29–32, Available from <http://www.ifi.unizh.ch/cl/yuste/postworkshop/postworkshop.htm>.
- Bowker, L. and J. Pearson. 2002. *Working with specialized language. A practical guide to using corpora*. London: Routledge.
- Department of Education and Training. 1993. *IsiZulu terminology and orthography No. 4*. Pretoria: The Government Printer.
- De Schryver, G. -M. and R. Gauton. 2002. The Zulu locative prefix ku- revisited: A corpus-based approach. *Southern African Linguistics and Applied Language Studies* 20 (4):201–220.
- Geldenhuis, J. D. U., E. E. Viljoen and K. D. Mboweni-Marais. 1995. *Trilingual business dictionary/Idikshinari yebhizinisi yezilimi ezintathu/Drietalinge Sakewoordeboek*. Pretoria: J. L. van Schaik.
- Johansson, S. 1998. On the role of corpora in cross-linguistic research. In *Corpora and cross-linguistic research: Theory, method, and case studies*, ed. S. Johansson and S. Oksefjell, 3–24, Amsterdam: Rodopi.
- Kumalo, M. B. 1995. *English/IsiZulu linguistic and literary terms/Amatemu esiNgisi/esizulu elingwistiksi nelithrisha*. Florida: Vivlia.
- Laviosa, S. 2003. Corpora and the translator. In *Computers and translation: A translator's guide*, ed. H. Somers. Amsterdam: John Benjamins.
- Maia, B. 2003. The pedagogical and linguistic research implications of the GC to on-line parallel and comparable corpora. In *CP3A – Copora Paralelos, Aplicações e Algoritmos Associados, Braga, 13 de Maio de 2003*, ed. J. J. Almeida, 31–32, Braga: Universidade do Minho. <http://www.linguateca.pt/>.

- McEnergy, A. and Z. Xiao. (forthcoming). Parallel and comparable corpora: What are they up to? In *Corpus linguistics and translation studies*, ed. G. James. Clevedon: Multilingual Matters. <http://www.lancs.ac.uk/postgrad/xiaoz/publications.htm>.
- Pearson, J. 1998. *Terms in context*. Amsterdam: John Benjamins.
- Rogers, T. 1999. *A basic guide to computer terms/n Basiese gids vir rekenaar terme/ Incwadi eyisisekelo samagama ekhompuyutha/ Hlahlo ya theo ka mareo a khomphuthara*. Pretoria: J. L. van Schaik.
- Scott, M. 1999. *WordSmith Tools version 3*. Oxford: Oxford University Press. See for this software also <http://www.lexically.net/wordsmith/index.html>.
- Sinclair, J. M. 1995. Corpus typology. A framework for classification. In *Studies in Anglistics. Acta Universitatis Stockholmienses*, ed. G. Melchers and B. Warren, 17–33, Stockholm: Almqvist & Wiksell International. Also available from <http://www.ilc.pi.cnr.it/EAGLES96/corpusyp/corpusyp.html>.
- Taljad, E. and G.-M. de Schryver. 2002. Semi-automatic term extraction for the African languages, with special reference to Northern Sotho, *Lexikos* 12 (AFRILEX series 12:44–74.)
- Uzar, R. and J. Walinski. 2000. A comparability toolkit: Some practical issues for terminology extraction. In *PALC '99: Practical applications in language corpora. Papers from the International Conference at the University of Lodz, 15–18 April 1999*, ed. B. Lewandowska-Tomaszczyk and P. J. Melia, 445–457, Lodz Studies in Language 1. Frankfurt am Main: Peter Lang.
- Zanettin, F. 1998. Bilingual comparable corpora and the training of translators. *Meta* 43 (4):616–630.
- . 2002a. DIY Corpora: The WWW and the translator. In *Training the language services provider for the new millennium*, ed. B. Maia, J. Haller and M. Urych, 239–248, Porto: Faculdade de Letras, Universidade do Porto.
- Zanettin, F. 2002b. Corpora in translation practice. In *Language resources for translation work and research, LREC 2002 Workshop Proceedings, Las Palmas de Gran Canaria*, ed. E. Yuste-Rodrigo, 10–14, <http://www.ifi.unizh.ch/cl/yuste/postworkshop/postworkshop.htm>.