

Analysing the English-Xhosa parallel corpus of technical texts with Paraconc: a case study of term formation processes

Koliswa Moropa

Department of African Languages, University of South Africa, PO Box 392, UNISA 0003

e-mail: moropck@unisa.ac.za

Abstract: The aim of this article is to show how corpus-based research, using computer tools can contribute to the development of technical terminology for Xhosa by translators. Paraconc has been used in analysing a parallel corpus of English-Xhosa texts. Paraconc is a parallel concordancer which makes it easy to analyse translated texts. This software program was developed by Michael Barlow (1995, 2003). He designed this software for linguists and translators who wish to work with translated texts. The alignment process, which is very important for the successful operation of this software, creates links between the source text and the target text. Some screenshots which show alignment of texts at sentence level, merging and splitting of sentences, word frequency lists, parallel concordances, etc. are cited. The author further demonstrates how parallel concordancing, and to a lesser extent, sentence alignment in Paraconc, have been used for identifying terms in the English-Xhosa Parallel Corpus.

Introduction

Since the dawn of democracy in South Africa, there has been a significant increase in the demand for translations of hegemonic languages into the previously marginalised languages and vice versa. Although a lot of translation is being done into African languages, the problem that faces us is that the translated texts, due to a lack of terminology, are not of a uniformly 'acceptable' standard. The aim of this article is therefore to illustrate how corpus-based research, using computer tools can contribute to the development of technical terminology for Xhosa by translators. The use of computers has made a corpus-based approach to the study of language extremely advantageous, as computers make the identification and analysis of complex patterns of language use possible. They also allow for the storage and analysis of much larger databases of natural language than could be dealt with by hand. Moreover, computers provide constant and reliable analyses. They can also be used interactively, allowing the human analyst to make difficult linguistic judgements while the computer takes care of record-keeping (Biber *et al.*, 1998: 4).

In the first section of this article, key concepts such as corpus, corpus-based research and uses of parallel corpora are outlined. The second section sketches the design of the English-Xhosa Parallel Corpus on which the research is based. In the third section, a brief illustration on how Paraconc is used in analysing parallel texts is provided. The fourth section focuses on how parallel concordancing and sentence alignment in Paraconc are used for identifying terms.

Corpus linguistics and corpus-based research

According to Laviosa-Braithwaite (1996: 14), corpus linguistics can be defined as a branch of general linguistics that involves the analysis of large machine-readable corpora of running text, using a variety of software tools designed specifically for this purpose. Corpus linguistics is characterised by its unique approach to the study of language, based on the integration of four important elements namely — data, description, theory and methodology. The data is compiled and designed according to a set of principles and it is examined by means of computer tools. The facts which are discovered about language are systematically organised in new descriptions of language

behaviour. Laviosa-Braithwaite (1996: 14) states that these descriptions 'feed into linguistic theory, where concepts and language models are created to explain and accommodate the phenomena empirically observed and hypotheses are put forward for further testing'.

Biber *et al.* (1998: 4) cite the following characteristics of corpus-based analysis:

- it is empirical, analysing the actual patterns of use in natural texts;
- it utilises a large and principled collection of natural texts, known as a corpus, as the basis for analysis;
- it makes extensive use of computers for analysis, using both automatic and interactive techniques;
- it depends on both quantitative and qualitative analytical techniques.

According to Biber *et al.* (1998: 11–12), one of the most important insights of corpus-based research is that intuitions which linguists have held frequently prove to be incorrect when they are tested empirically against actual patterns in large text corpora. Another insight is the realisation that few linguistic descriptions are adequate for a language as a whole. Empirical analyses of large corpora repeatedly show that there are important systemic differences among registers at all linguistic levels.

One of the strengths of the corpus-based approach is that it can be applied to empirical investigations in almost any area of linguistics. For example, within sociolinguistics, corpus-based techniques allow investigations of dialect and register patterns that previously could not be addressed. Studies of style are also possible with corpus-based approaches. Individual authors or styles across historical periods can be investigated in a more comprehensive way than in the past. Corpus-based studies are also applicable to educational linguistics. The results of large-scale studies are helpful in designing effective materials and activities for classroom and workplace training, allowing educators to help students with the language that is actually used in different target settings. Within educational linguistics, the field of language testing, too, can benefit from results of corpus-based studies, designing tests which conform to the actual language that students will be using on a regular basis. Corpus-based studies also spilled over into the area of translation studies in the early 1990s.

Corpus-based translation research

Mona Baker of Manchester University in the UK pioneered corpus-based translation research. According to Baker (1993), corpora have a profound effect on translation studies as they enable researchers to identify features of translated texts that will help us understand what translation is and how it works. Originally, the term corpus meant any collection of writings in a processed or unprocessed form by a specific author. Baker (1995) states that the word corpus has often been used in translation studies to refer to fairly small collections of text which are not held in electronic form and which are therefore analysed manually. According to Baker (1995: 225), with the growth of corpus linguistics, the definition has changed in three important ways:

- corpus now means primarily a collection of texts held in machine-readable form and capable of being analysed automatically or semi-automatically in a variety of ways;
- a corpus is no longer restricted to written texts but can include spoken texts;
- a corpus may include a number of texts from a variety of sources by many writers and speakers and on a multitude of topics. What is important is that it is put together for a particular purpose and according to explicit design criteria in order to ensure that it is representative of the given area or sample of language for which it aims to account.

Sinclair (1995: 17) mentions that computer-held corpora have been in existence for nearly thirty years, but the terminology for describing and classifying corpora has not been settled. In his classification, Sinclair distinguishes various types such as reference, monitor, parallel, comparable corpora, etc. Since this article is based on a parallel corpus, it is imperative to define this type of corpus.

Parallel corpora

According to Sinclair (1995: 32), a parallel corpus is a 'collection of texts, each of which is translated into one or more other languages than the original'. Sinclair observes that parallel corpora are an object of interest at present as they offer the opportunity to align original texts and their translations and gain insights into the nature of translation. According to Baker (1995), the most important contribution of parallel corpora is that they support a shift of emphasis from prescription to description. They allow researchers to establish, in an unbiased manner, how translators overcome difficulties in translation practice, using this proof to provide a practical example in the training of translators. Baker assumes that parallel corpora will soon become vital in material writing, computer-aided training and in improving the performance of machine translation systems. Kenny (1997: 387) defines parallel corpora as 'structured electronic collections of original texts in one language and their translations into one or several other languages'. Parallel corpora can be bilingual when they comprise original texts and their translated versions, or multilingual where corpora contain translations into several target languages of the same source language.

Using parallel corpora in translation studies

Bowker and Pearson (2002) argue that there are three groups — namely, the teachers/students of translation, language learners and computational linguists who are interested in using parallel corpora. According to Bowker and Pearson (2002: 103), a teacher/student of translation can use parallel corpora in the following ways:

- to examine how translators have handled certain linguistic features;
- to examine how cohesive devices have been translated;
- to look more closely at what happens to culture-specific references in translation;
- to examine what has not been translated and think of the reasons for this; and
- to look at what has been added to a translation.

The above scholars further state that, a language learner may use parallel corpus as follows:

- in the same way one would use a bilingual dictionary;
- to establish whether a particular translation found in the dictionary is actually used;
- to establish how words are used, i.e. whether a particular word favours a certain syntactic pattern or prefers a certain group of adjectives; and
- to find out how to present something in another language.

Bowker and Pearson (2002: 95) indicate that computational linguists differ from the other two groups. They are not interested in the usefulness of parallel corpora for translation purposes or for language learning. Computational linguists use parallel corpora as a testing ground for developing alignment software. They want to establish whether or not the characteristics of texts can be expressed computationally so as to facilitate the development of alignment programmes.

Parallel corpora have already been used by a number of researchers in translation. Several case studies are summarised below.

King (1997: 393) reports on the parallel concordancing project which was started in 1993 and which was funded by the European Union. Originally, the project involved ten tertiary institutions in six EU member countries. Although the multilingual parallel corpus is in the languages of the six original participating countries (Danish, English, French, German, Greek, Italian), the software developed allows for the processing of many other languages. Other languages that were added with effect from 1997 were Finnish, Portuguese, Spanish and Swedish. The user can select any pair of languages. The size of the corpus is 3 134 000 words at the present moment and King describes this capacity as small but respectable. A single search can be made in one fraction of this corpus at a time, as it consists of texts spread across six languages. King states that a parallel corpus can be used to:

- examine translator behaviour in a systematic and principled way;
- check the behaviour of a translator against a bilingual dictionary; and
- examine claims made in theoretical translation terms.

As regards theoretical issues on translation, King (1997) examined one of the translation universals defined by Baker (1993: 244) i.e. 'a tendency to avoid repetitions which occur in source texts, either by omitting them or rewording them'. In investigating this universal feature of translation, King selected a Greek original novel and its translated English version. King's (1997) study showed that parallel corpora can be used in investigating universal features of translation. King emphasizes the fact that the research he conducted is just an exploration intended to demonstrate how a parallel corpus stored on a computer can be used.

Pearson (2003: 19), who believes that parallel corpora have a role to play in the training of translators, used such corpora to investigate the translation of culture-specific information. Here, by culture-specific information she refers to information about events, institutions, organisations or individuals in the source culture. The parallel corpus showed that a number of varying solutions were adopted. For example, in examining the rendering of references in translation, it was found that in some target texts the names of individual researchers were often omitted. Pearson says that the aim is not to suggest what students should do, but by looking at how professional translators have solved the problems, student translators at least obtain some indication of what to do. Students will also see for themselves what actually happens in translation, that different solutions may be appropriate for different situations and that one culture-specific reference can be treated in a number of different ways.

With regard to research in the African languages a number of scholars are using parallel corpora. Madiba (2004), using the Special Language Corpora for African Languages (SPeL CAL), illustrates how parallel corpora can be used as tools for developing the indigenous languages of South Africa. The SPeL CAL project was born out of the need for language resources to support the implementation of South Africa's multilingual language policy adopted after the democratic changes of 1994. The purpose of SPeL CAL is twofold:

- To provide a language resource for the compilation of specialised dictionaries, terminology lists and glossaries in the official African languages of South Africa;
- To provide a resource for research in linguistic fields such as terminology, terminography, translation, language for special purposes (LSP) and second language teaching (SLT) (Madiba, 2004: 136).

In a pilot analysis, Madiba (2004) used Multiconcord to analyse translation equivalents of terms like 'act', 'legislation', 'rule', 'order' and 'law' in a parallel corpus of English-Venda texts of *The Constitution of the Republic of South Africa* (Republic of South Africa 1996).

Gauton and De Schryver (2004) demonstrate how special-purpose multilingual and parallel corpora can be used as a translator's tool in finding suitable term equivalents when translating technical texts from English into Zulu. In conducting the research, the following general-language corpora were used:

- The University of Pretoria Zulu Corpus (PZC), an electronic corpus of five million running Zulu words established at the University of Pretoria by De Schryver and Dlomo. The corpus comprises literary texts, religious texts, internet files and pamphlets (Case study one)
- The University of Pretoria Internet English Corpus (PIEC), an electronic corpus of 12.4 million English words 'culled' from the Internet by Gauton (Case study two)

In Case study one, multilingual corpora were used to investigate terminology in the translation of HIV/AIDS texts. Multilingual corpora comprised a Zulu corpus of 7698 words and an English corpus of 10 426 words. HIV/AIDS terminology was identified in both corpora by making use of the Key Words function of WordSmith Tools. In Case study two, parallel corpora dealing with labour issues were used to investigate labour terminology in order to determine the usefulness of such corpora as a resource for the translation of technical texts into Zulu. The latter corpus comprised an English text of 1 479 words and its Zulu translated version of 1 004 running words, and was queried by means of Paraconc.

With regard to the methodologies, the best terminological results were obtained when using parallel corpora. Gauton and De Schryver (2004) mention that when a translator has access to parallel corpora in a special field, such corpora can be loaded into parallel concordancing software

and translation equivalents can be easily identified with or without sentence alignment. In such cases, parallel corpora function as translation memory.

The English-Xhosa Parallel Corpus

The compilation of the parallel corpus of English-Xhosa texts, was done in consultation with the National Language Service (NLS), a division of the Department of Arts and Culture, formerly known as the Department of Arts, Culture, Science and Technology (DACST). The English-Xhosa Parallel Corpus is composed of the following English texts and their translated Xhosa versions:

- The 1997 Annual Report of the Department of Arts, Culture, Science and Technology (DACST)
- A Short Guide to the White Paper on Local Government (1998)
- 2001/2002 Pan South African Language Board (PanSALB) Annual Report
- Three manuals on the Promotion of Access to Information Act (PAIA manuals)

The above texts that constitute the English-Xhosa Parallel Corpus can be regarded as authentic texts because they have not been created specifically for the purposes of this research. According to Bowker and Pearson (2002: 9), an authentic text is a natural text that consists of genuine communication between people under normal circumstances and which is an example of real 'live' language usage. It is not a text that has been created for the purpose of being included in a corpus in order to demonstrate a particular point of grammar, for example.

Domains of individual texts

The selected texts comprise annual reports and other informative documents that were translated between 1998 and 2003.

The 1997 Annual Report of the Department of Arts, Culture, Science and Technology (DACST)

The Department of Arts, Culture, Science and Technology was formerly divided into two major sections, i.e. Arts and Culture, and Science and Technology. The DACST Annual Report focuses on communication, management services, financial management, national archives of South Africa, language services, arts and culture management, science and technology.

A Short Guide to the White Paper on Local Government

A Short Guide to the White Paper on Local Government is a brief introduction to the White Paper on local government. The White Paper contains the government's vision of a new local government system for South Africa and how to achieve this vision. This guide does not cover everything that is in the White Paper. It only summarises the key issues. It starts by defining the white paper and then describes integrated developmental planning, local government training, administration systems, corporatisation and municipal finances.

The 2001/2002 Annual Report of the Pan South African Language Board (PanSALB)

The Pan South African Language Board is a body that has to provide for the recognition, promotion and development of all the South African languages, with particular attention being given to those that were previously marginalised. In order to achieve its goals, the Board has formed structures to facilitate the achievement of these goals — namely, National Lexicography Units (NLUs), National Language Bodies (NLBs) and Provincial Language Committees (PLCs). The focus areas of the annual report are lexicography and terminology, linguistic human rights and advocacy, development of languages, research and development and the financial statements.

Three manuals on the Promotion of Access to Information Act

The advent of democracy in South Africa brought a new ethos to the society such as transparency, accountability and citizenry. Access to information is an entrenched right which enables citizens to make informed choices, thereby enabling them to access services rendered by government

departments. The three manuals on the Promotion of Access to Information Act, reinforce the confidentiality of the information by stipulating that, in terms of the Act, information may only be disclosed to the person to whom it relates or to that person's authorised representative.

All the selected texts deal with aspects of administration, management, information technology and finances.

Publication dates

The Annual Report of the Department of Arts, Culture, Science and Technology of 1997 was translated and published in 1998. A Short Guide to the White Paper on Local Government was translated and published in 1998 and it is also found on the government website: www.dplg.gov.za. The Pan South African Language Board (PanSALB) Annual Report of 2001/2002 was translated in 2002. The three manuals on the Promotion of Access to Information Act; namely the South African Maritime Safety Authority (SAMSA), South African Revenue Services (SARS) and the manual to Independent Complaints Directorate (ICD) were translated in 2003.

Translators of the texts

The DACST Annual Report was translated by four translators, namely Fumi Dyubhele, Zandile Hadebe, Koliswa Moropa and Hlumela Sondlo, lecturers in the Department of African languages at the University of South Africa at that time. The translator of A Short Guide to the White Paper on Local Government is not acknowledged in the publication and is therefore unknown. The Annual Report of PanSALB (2001/2002) was translated by Zama Bekeweni, a renowned professional Xhosa translator, who was one of the compilers of the *Eskom glossary of energy terms* (Corporate Communications Department 2000). The manuals on promotion of access to information were translated by Bulelwa Nokele and Thembela Sineke who are practising translators in the Eastern Cape.

Size of corpus

The English-Xhosa Parallel Corpus comprises 82 163 English words and 62 538 Xhosa words. The difference in the number of words between the two sub-corpora of English and Xhosa is to be expected because Xhosa is written conjunctively. What is in ten words in English is written in five words in Xhosa, for example.

The following example from the PanSALB Annual Report (2001/2002) illustrates this:

ST: There is a need to improve the system of finance. [10 words]

TT: *Kukho intswelo yokuphucula inkqubo yemali.* [5 words]

Figure 1 illustrates the structure of the English-Xhosa corpus. In the next section a short description of Paraconc is given which is followed by a discussion of how Paraconc is used in analysing parallel texts. Screenshots are cited from the English-Xhosa Parallel Corpus.

Paraconc — a parallel concordancer

Paraconc is a parallel concordancer which was developed by Michael Barlow (1995, 2003). Barlow designed Paraconc for linguists and translators who wish to work with translated texts. Paraconc, like other concordance programs, provides the following features:

- A concordancer, which finds and displays, in an easy-to-read format, all occurrences of a particular search term (and minor variations thereof);
- A collocation viewer, which allows users to see which words go together; frequency lists, etc.
- Paraconc can also be used by teachers and other researchers who are interested in analysing multilingual texts. The alignment process is very important for the successful operation of the software. Kraif (2002: 275) describes alignment as involving a parallel segmentation of both texts into smaller logical units such as paragraphs, sentences or even phrases in such a way that the n^{th} segment of source text and n^{th} segment of target text are mutual translations. Paraconc is a software programme that makes it easy to analyse translated texts.

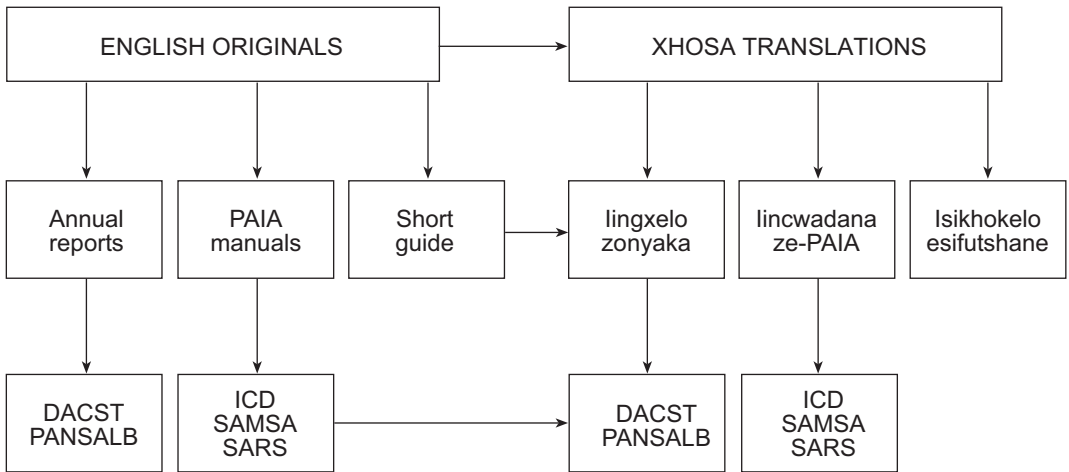


Figure 1: The structure of the English-Xhosa Parallel Corpus

How Paraconc functions

After the texts had been saved, they can be loaded and saved in a workspace created in Paraconc. Figure 2, is an illustration of parallel texts such as the PanSALB Annual Report, SARS Manual, the DACST Annual Report, etc., which have been loaded in a workspace in Paraconc. Since Xhosa does not appear in the list of languages displayed in the right-hand dialogue box, 'Afrikaans' was used as a label. There is no limit to the size of the texts that can be loaded. Creating a workspace in Paraconc saves the researcher from reloading the corpus each time he/she wants to use it.

Alignment of texts

The alignment process is very important for the successful operation of the software. In the alignment process, the texts are matched at sentence level so that a sentence in the source text finds a corresponding sentence in the target text. In aligning the texts, the sentence is used as the basic alignment segment. However, this does not mean that each sentence in Language A (English) is aligned with exactly one sentence of Language B (Xhosa) since sometimes a sentence in English is equivalent to two sentences in Xhosa or may sometimes be omitted altogether. When an original sentence has only one corresponding sentence in the translated text, we have a 1:1 correspondence as shown in Figure 3.

If the sentences are not aligned properly, (which often happens), a menu of options allows the splitting and merging of sentences or segments (cf. drop-down menu on the right in Figure 4).

Once the texts have been aligned, a word list of two parallel texts can be created. The corpus frequency commands create a word list for the two parallel texts and the results are displayed in two parallel windows. The word list may be arranged according to the order of frequency (cf. Figure 5) or it can be arranged alphabetically (cf. Figure 6). In Figure 5 we notice that, in English, the word with the highest frequency is 'the' with 398 occurrences. In Xhosa, the word with highest frequency is a conjunctive *ukuba* 'that' or 'if' with 75 occurrences. The word list gives an overall idea of what information can be searched from the text.

In Figure 6, the word list is arranged alphabetically. All the Xhosa words that appear in the list start with *aba* which denotes various concords and prefixes. For example, the first *aba-* is a demonstrative, the second *aba-* in *ababekade* is an adverbial concord and the third *aba-* in *ababoneleli* is a Class 2 noun prefix.

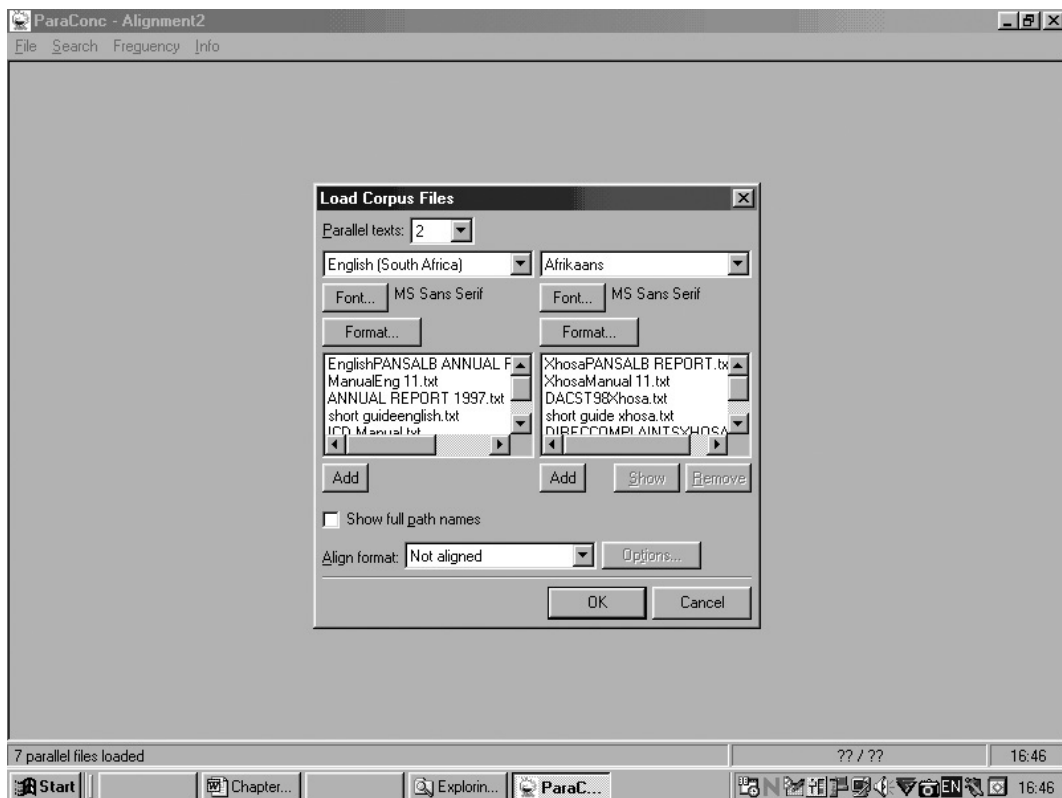


Figure 2: Parallel texts loaded in workspace

Parallel concordancing

One of the main advantages of a parallel corpus is that it enables one to access a particular word in context in both the source language and the target language. With ParaConc, searching for a word in one language gives concordances in both the search language (English in this case in the upper window) and the parallel language (Xhosa in the lower window). The search term is centred within the concordance window and is displayed in Key Word In Context (KWIC) format (cf. parallel concordance of local government/*ulawulozidolophu* in Figure 7).

In the next section, the translation of each term is shown in a parallel concordance line. Parallel concordances show exactly where the example is found in the text.

Parallel concordancing for identifying terms

From the concordances which were obtained from the English-Xhosa Parallel Corpus, it became clear that the Xhosa translators employed the following term creation strategies, i.e. using a loan word, using a borrowed synonym, compounding, derivation and paraphrasing. The majority of terms which are discussed below are found in most texts that constitute the corpus.

Using a loan word

Since no group lives in isolation of other groups, the contact between groups leads to the exchange of concepts and their words. In his terminology guidelines for translators, Trew (1994: 96) says that translators should not avoid loan words that are widely understood, or loan words that will express the original meaning precisely to the target reader. What they should avoid is using a loan word in a way

A Short Guide to the White Paper on the Local Government	ISIKHOKELO ESIFUTSHANE SEPHEPHA LENGCACISONKQUBO YOLAWULOZIDOLOPHU
This guide is a short introduction to the White Paper on Local Government, first published in March 1998.	Esi sikhokelo yintshayelelo emfutshane yePhepha leNgcacisonkqubo yoLawulozidolophu, esipapashiwe okokuqala ngoMatshu 1998.
The White Paper contains the government's vision of a new local government system for South Africa and how to achieve this vision.	IPhepha leNgcacisonkqubo liqulethe imbono karhulumente yenkqubo yolawulo olutsha lweedolophu emZantsi Afrika kwakunye nendlela yokufikelela le mbono.
This guide does not cover everything that is in the White Paper.	Esi sikhokelo asiqukanga yonke into efumaneka kwiPhepha leNgcacisonkqubo.
It only summarises the most important matters - the key issues.	Eyona nto siyenzayo kukushwankathela kuphela - inuba engundoqo
Introduction	INTSHAYELELO
What is the White Paper on Local Government?	YINTONI IPHEPHA LENGCACISONKQUBO YOLAWULOZIDOLOPHU?
The present local government system is a temporary (transitional) system.	Inkqubo yangoku yolawulozidolophu yeyethutyana (eyexesha lenguqo).
It is intended to take us from the old apartheid system to a new, democratic system.	Injongo yayo kukusisusa kwinkqubo endala yocalucalulo ukusisa kwentsha eyedemokhrasi.
It is the government's task to ensure that a new system is in place by the time the next local government elections take place in 1999.	Ngumsebenzi karhulumente ukuqinisekisa ukuba inkqubo entsha iyasebenza ngexesha lonjalo olulandelayo lolawulozidolophu ngowe-1999.
The White Paper is the plan for the new system of local government.	IPhepha leNgcacisonkqubo sisicwangciso senkqubo entsha yolawulo-zidolophu.
It was developed by the Ministry of Provincial Affairs and Constitutional Development after months of research and discussions with everyone concerned.	Yaphuhliswa yiOfisi yoBuphathiswa bemiCinbi yemaPhondo noPhuhliso lomThethosisekelo emva kweenyanga zokuphanda nokuxoxisana naye wonke umntu ochaphazelekayo.
It says what kind of local government system we want, and how we can achieve it.	Ixela uhlobo lweenkqubo zolawulo-zidolophu esizifunayo, nendlela esonokuthi sizifikelele ngayo.
Of course, it is not enough just to have a plan.	Phofu ke, akulingananga ukuba nesicwangciso nje kuphela.
Further steps will have to be taken to put the plan into action.	Kufuneka ukuthatyathwe amanye amanyathelo ukwenza ukuba isicwangciso sisebenze.
Laws will have to be passed, new boundaries will have to be demarcated, new programmes will have to be developed, and people will have to be trained in new concepts for local government.	Kuya kufuneka kudluliswe imithetho, kusikwe inuda emitsha, kuphuhliswe iinkqubo ezintsha, yaye abantu kufuneka baqeqeshwe ngeendlela ezintsha

Figure 3: Alignments of texts at sentence level

that creates obstacles to understanding. According to Cluver (1989: 267–268), some of the conditions which are necessary for borrowing to take place are:

- Contact: Two speech communities must be in regular contact for borrowing to occur (in this case English and Xhosa speakers)
- Linguistic similarity: Some linguistic similarity seems necessary for borrowing to take place
- Language attitudes: The attitude of the receiving community can play an important role in determining the number of loan words and where their use may be permitted
- Cultural and technological inequality: The receiving language must have a need for the loan words. The receiving language usually has no technical equivalents and it therefore takes over the imported technology and its terminology from the donating culture

Cluver (1989: 270) says that borrowing is an important way in which technical languages expand their vocabularies, and this leads to an internationally accepted terminology which makes technical communication across language boundaries fairly easy.

In areas where two languages are unequal, as in this case, English and Xhosa, the less developed technical language, Xhosa, has taken over concepts and terms from English. Using a loan word is not new in Xhosa. For instance, when Xhosa was reduced into writing, it had to receive items which were foreign to the Xhosa culture — from English, for example Satyo (1981: 54) affirms this:

Sinokuboleka igama kwezinye iilwimi. Cinga ke ngamagama afana nala: imoto, itreyini, isikolo, itshizili, ikani, ibrukhwe, ihempe njalo njalo.

[We can borrow a word from other languages. Think of words such as these: *imoto*, 'motor car', *itreyini* 'train', *isikolo* 'school', *itshizili* 'chisel', *ikani* 'a can', *ibrukhwe* 'broek', *ihempe* 'hemp' etc.]

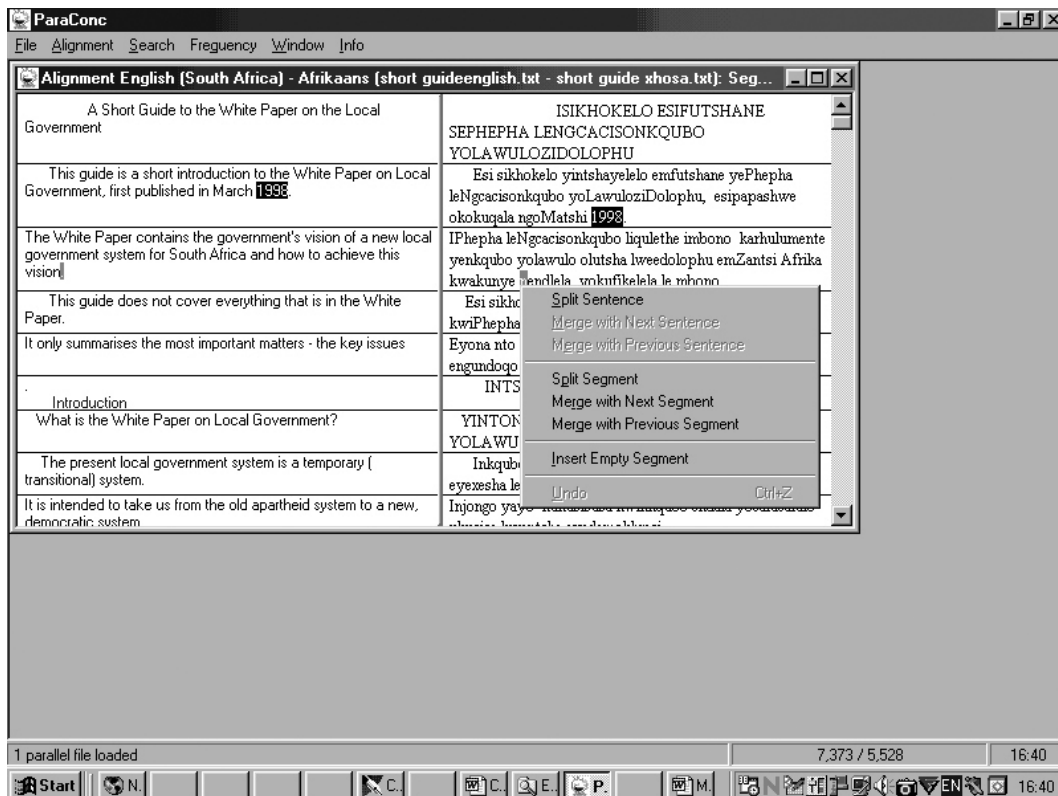


Figure 4: Merging and splitting of segments/sentences

It was found that in the parallel corpus of English-Xhosa texts, the translators used the following loan words:

- An indigenised loan word
- An indigenised loan word preceded by explanation
- A pure loan word
- A pure loan word preceded by an explanation

These term formation processes are now discussed.

An indigenised loan word

According to Wallmach and Kruger (1999: 281), indigenising a loan word means ‘modifying the word slightly to remove some of the ‘foreignness’ of the word and spelling it according to the orthography of the language which is borrowing the word’. One notices that Xhosa is capable of creating new words and expressions to accommodate new concepts and cultural items and, to an even a greater extent to absorb and ‘xhosaise foreign words and terms’ (Pahl *et al.*, 1989: 32).

An indigenised loan word preceded by explanation

In Table 2, the translator used indigenised loan words preceded by an explanation and the loan word is put in brackets. The translators provided an explanation because terms such as ‘hardware’ and ‘software’ are computer terms which may be new to the target reader. The translators provided an explanation so that the loan words would not create ‘an obstacle to understanding’ (Trew 1994: 96).

English (South Africa)			Afrikaans		
Count	Pct	Word	Count	Pct	Word
398	5.3981%	the	77	1.3929%	ukuba
308	4.1774%	to	74	1.3386%	oomasipala
291	3.9468%	and	71	1.2844%	kufuneka
223	3.0245%	of	62	1.1216%	-
146	1.9802%	a	45	0.8140%	yaye
129	1.7496%	government	41	0.7417%	kwakunye
120	1.6276%	for	38	0.6874%	oku
107	1.4512%	will	30	0.5427%	kamasipala
103	1.3970%	local	27	0.4884%	inkqubo
101	1.3699%	municipalities	26	0.4703%	kuya
99	1.3427%	in	25	0.4522%	lengcacisonkqubo
83	1.1257%	is	22	0.3980%	okanye
83	1.1257%	that	22	0.3980%	wesizwe
78	1.0579%	be	21	0.3799%	amagunya
75	1.0172%	are	20	0.3618%	ibhunga
58	0.7867%	municipal	20	0.3618%	iinkonzo
57	0.7731%	this	20	0.3618%	zesiseko
56	0.7595%	development	19	0.3437%	imisebenzi
55	0.7460%	-	19	0.3437%	ulawulozidolophu
54	0.7324%	with	18	0.3256%	icandelo
52	0.7053%	have	18	0.3256%	imali
48	0.6510%	national	17	0.3075%	izinto
45	0.6103%	services	17	0.3075%	phakathi
43	0.5832%	can	17	0.3075%	ukwenza
41	0.5561%	they	17	0.3075%	umasipala
39	0.5290%	on	17	0.3075%	yesisizwe

Figure 5: Word list from Short Guide (1998) in order of frequency

The translators did not rely on the bilingual dictionary, for example, translations of words like brochure (Table 1), hardware, software (Table 2), are found in the Fischer *et al.* *English-Xhosa Dictionary* (1992), but the translators created their own terms. One reason which led the translators to do this, is that some of the equivalents supplied are not appropriate to the given context and, where explanations are given, they tend to be very long. To cite only a few examples, in the Fischer *et al.* bilingual dictionary the translation for brochure is given as *incwadana* 'a booklet', hardware is described as *izixhobo zokulwa* 'weapons for fighting' and software is defined as *amanani, uluhlu lwenkqubo njl. njl. ezifakwa kwikompyuta* 'numbers, list of programmes etc., inserted into a computer'.

A pure loan word

Pure loan words are source language words which remain the same in the target text. Translators often use this strategy to deal with culture-specific items, modern concepts and buzz words. Acronyms and abbreviations are retained in their original form in the target text, but the necessary prefixes should be added in Xhosa.

Pure loan words are also used in translating abbreviations and acronyms. Abbreviations are always pronounced as a sequence of letters, and they function as normal word forms taking plural suffixes as well. It is not possible to translate abbreviations in Xhosa because the language is written conjunctively. For example, NLU stands for National Lexicography Unit, which is translated in the parallel corpus as *iQumrhu lesiZwe lokubhalwa kwesiChazi-magama*. In a literal sense, the translation reads: Unit of nation of lexicography.

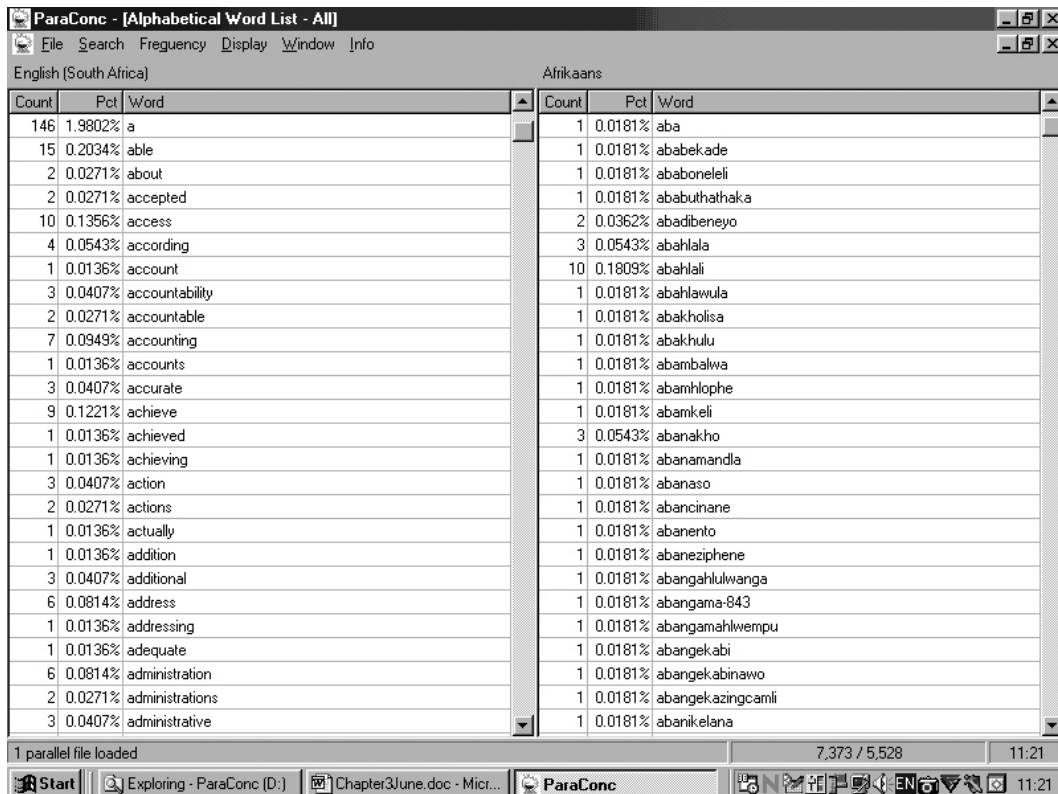


Figure 6: Word list arranged alphabetically from Short Guide (1998)

An acronym is formed by using one or more initial letters of an expression and forming a new word with them, for example, AIDS (Acquired Immune Deficiency Syndrome). The more widely used the acronyms are, the more likely they are to be assimilated into the language. Acronyms are assimilated into Xhosa and are written and pronounced as independent words, but the necessary prefixes should be added.

Just like abbreviations, it is not possible at this stage for Xhosa to form its own acronyms because it is written conjunctively.

A pure loan word preceded by explanation

In the following examples, the translator explained the loan word first and used a pure loan word in brackets.

As illustrated in Tables 4 and 6, the translators explained the terms ‘hardware’ and ‘software’ first and then provided a loan word in brackets. The explanations given by both translators convey the same meaning, for example, ‘hardware’ is translated as *izixhobo ezenza ikhompyutha ozibona ngamehlo* (parts that form a computer which you see) or *izixhobo eziphathekayo zekompyuta* (parts of the computer that you can touch), and ‘software’ is translated as *iinkqubo eziqulethwe yikhompyutha* (programmes contained in the computer) or *ezingaphathekiyo* (parts of the computer that you cannot touch). The only difference between the two translators is that after giving the explanation, the translator of the SAMSA Manual provided an indigenised loan word in brackets, whilst the translator of the PanSALB Annual Report gave a pure loan word.

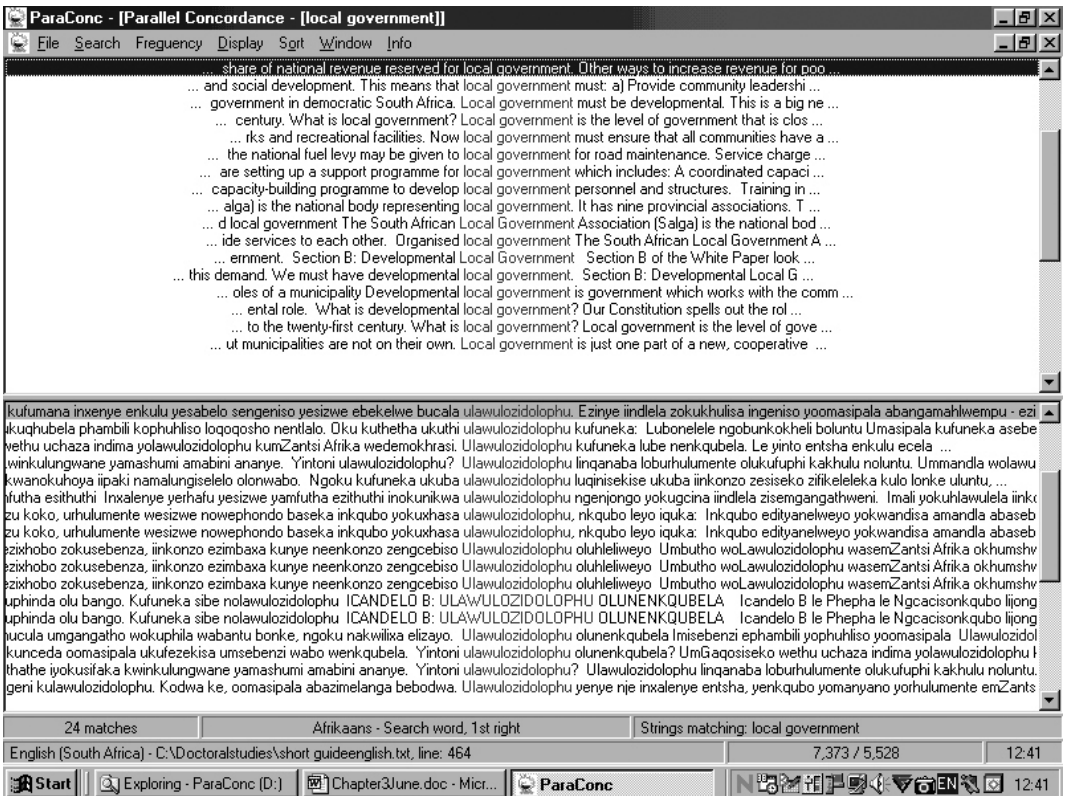


Figure 7: Parallel concordance list of local government > *ulawulozidolophu*

In the parallel corpus of English-Xhosa texts the translators used loan words for modern computer terms, as well as for financial and other technical terms. The translators probably opted for loan words because most of these items are new in the Xhosa culture. Xhosa uses many English loan words for foreign concepts, but this does not mean that Xhosa is unable to derive terms from its own vocabulary.

Using a borrowed synonym

In Xhosa, many loan words occur and are fully adapted to the system of the language. Sometimes a loan word is taken over from English or Afrikaans when a Xhosa word for the object already exists. Accordingly, the loan word and the Xhosa word co-occur in the lexicon and are synonyms, in that they have the same meaning. Some translators have used borrowed synonyms as shown in Table 7.

Compounding

Cluver (1989: 279) claims that compounding is probably the most important mechanism for creating new technical terms in any language. Compounding involves combining two or more words (two nouns, a noun and a verb, etc.) into one unit. The meaning of compounds is often dependent on the context in which they are used and the underlying relations between the constituting elements. In creating financial terms the translators used the noun *imali* (money) to depict the financial context as illustrated in Table 8, for example, *ubalo-mali*, *ucalulo-mali*, *ulwabiwo-mali*, *uhlahlolwabiwo-mali*, *utyalo-mali*, *inkxaso-mali*, etc.

Table 1: Indigenised loan words (for additional examples see Addendum 1)**1. account > iakhawunti SARS Manual (2003)**

b class=normal1> Line: 283

These records include SARS' financial records on own **[[account]]**

<b class=normal2> Line: N/A

Ezi ngxelo zigciniweyo ziquka iingxelo ze-SARS zezemali nge-**[[akhawunti]]**

2. brochure/s > iibhrowutsha SAMSA Manual (2003)

b class=normal1> Line: 251

... * Advisory and promotional pamphlets/**[[brochures]]**/posters*

<b class=normal2> Line: N/A

*Amaphetshana/**[[iibhrowutsha]]** iipowusta zengcebiso nezokwazisa*

3. cheque > itshekhi SARS Manual (2003)

b class=normal1> Line: 333

at a SARS cash office or by **[[cheque]]** ...

<b class=normal2> Line: N/A

kwi-ofisi ye-SARS yemali esesandleni okanye nge**[[tshekhi]]**

4. company > inkampani SARS Manual (2003)

<b class=normal1> Line: 700

... . Full names and surname: Identity/**[[company]]** number: D.

<b class=normal2> Line: N/A

Amagama azeleyo nefani: Inombolo yesazisi/ ye**[[nkampani]]**: D.

5. data > idata PanSALB (2001/2002)

<b class=normal1> Line: 185

(iii) Linguistic **[[data]]** and the presentation thereof in dictionaries...

<b class=normal2> Line: N/A

(iii) **[[idata]]** yolwimi nokuhlelwa kwayo kwisichazi-magama

6. democracy > idemokhrasi

<b class=normal1> Line: 245

Promote local **[[democracy]]** by facilitating the participation of citizens

<b class=normal2> Line: 491

Kukhuthazwa **[[idemokhrasi]]** yendawo ngokulungiselela ukuba abemi

7. deposit > idiphozithi SARS Manual (2003)

<b class=normal1> Line: 91

... processing your request after you have paid a **[[deposit]]**.

<b class=normal2> Line: N/A

... isisebenze isicelo sakho emva kokuba uhlawule **[[idiphozithi]]**.

The translator of the PanSALB Annual Report (2001/2002) used compounding in translating 'policy', 'strategy' and 'workshop', which were translated by a more general word *inkqubo* or *umgaqo* by most translators. The translator formed new terms by compounding noun + noun, for example:

strategy > *isicwangciso-qhinga* [*isicwangciso* 'plan' + *qhinga* 'cunning trick']

policy > *umgaqo-nkqubo* [*umgaqo* 'method' + *nkqubo* 'procedure'] and **workshop** > *uqeqesho-msebenzi* [*uqeqesho* 'training' + *umsebenzi* 'work']

Given the meaning of strategy as 'the skilful planning or management of anything', policy as 'a plan of action adopted or pursued by an individual, government, party or business' and workshop as 'engagement of a group of people on a creative project or subject', the compound nouns seem to be appropriate translations.

Derivation

There are two word formation processes that involve adding an affix to a root namely, prefixation and suffixation. Suffixes are morphemes added after another morpheme and prefixes are added to the beginning of another morpheme.

Table 2: Indigenised loan word preceded by explanation**SAMSA Manual (2003)****1. cash > (i)mali esesandleni (ikheshi)**

b class=normal1> Line: 333

may pay the fees either in **[[cash]]** at a SARS cash office or by cheque <b class=normal2> Line: N/A
imirhumo nge**[[mali esesandleni (ikheshi)]]** kwi-ofisi ye-SARS yemali

2. hardware > izixhobo ezenza ikhompuyutha ozibona ngamehlo (ihadiwe)

b class=normal1> Line: 162

... * recovery * **[[Hardware]]** * Software * Internet * Procedures ...

<b class=normal2> Line: 162

* konakele * **[[Izixhobo ezenza ikhompuyutha ozibona ngamehlo (ihadiwe)]]** *

3. schedule > uludwe lwenkqubo (ishedyuli)

<b class=normal1> Line: 150

... , 1996, various other Acts listed in **[[Schedule]]** 1 to the South African

<b class=normal2> Line: N/A

1996, neminye iMithetho edweliswe ku**[[Ludwe lweNkqubo (iShedyuli)]]**

4. software > iinkqubo eziqulethwe yikhompuyutha (isoftiwe)

<b class=normal1> Line: 162

* Hardware * **Software** * Internet *

<b class=normal2> Line: 162

* Ihadiwe * **[[linkqubo eziqulethwe yikhompuyutha (isoftiwe)]]** * I-intanethi*.

Table 3: Modern concepts**SAMSA (Manual 2003)****1. e-mail > e-mail**

b class=normal1> Line: 220

... **[[fax]]** number or **[[e-mail]]**address set out in section 2 ...

<b class=normal2> Line: N/A

...kwinombolo yefeksi okanye kwidilesi ye-**[[e-mail]]** ebhalwe kwicandelo

Table 4: Abbreviations**PanSALB 2001/2002****1. NLUs > NLUs**

b class=normal1> Line: 241

... guidance in the compilation of **[[corpora]]** and dictionaries for the **[[NLUs]]**

<b class=normal2> Line: N/A

Uqeqesho nokukhokelwa kwee**[[NLUs]]** kwiinkqubo zokuqokelelwa ...

SARS Manual (2003)**2. EU > EU**

<b class=normal1> Line: 204

... Agreement between South Africa and the **[[EU]]**, ...

<b class=normal2> Line: N/A

* noVumelwano lwentsebenziswano phakathi koMzantsi Afrika ne-**[[EU]]**, ...

Table 5: Acronyms**SARS Manual (2003)****1. SARS > SARS**

<b class=normal1> Line: 307

... to request relates to information about **[[SARS]]**

<b class=normal2> Line: 303

... onqwenela ukuyicela inxulumene na nolwazi olunge-**[[SARS]]**

2. VAT > VAT

<b class=normal1> Line: 205

... ied by SARS branch offices dealing with **[[VAT]]**. VAT refunds in respect

<b class=normal2> Line: 202

... qiniselwe zii-ofisi zesebe le-SARS ezijongene ne-**[[VAT]]**. limbuyekezo ze-

Table 6: Pure loan words preceded by explanation**PanSALB (2001/2002)****1. corpora > ingqokelela yesigama (corpora)**

<b class=normal1> Line: 196

... · Compiling and using **[[corpora]]**

<b class=normal2> Line: N/A

Ukubhala isichazi-magama ngokusebenzisa **[[ingqokelela yesigama (corpora)]]**

2. hardware > izixhobo eziphathekayo zekompyuta (hardware)

<b class=normal1> Line: 213

... assessments with regard to computer **[[hardware]]**

<b class=normal2> Line: N/A

... iimfuno ezimalunga ne**[[zixhobo eziphathekayo zekompyuta (hardware)]]**

3. software > ezingaphathekiyo (software)

<b class=normal1> Line: 213

... to computer hardware and **[[software]]** were carried out in all

<b class=normal2> Line: N/A

... zekompyuta (hardware) kwanezo **[[zingaphathekiyo (software)]]**

Prefixation and suffixation

Prefixation and suffixation offer another method that has been used by the translators to create terms. In Xhosa, when a noun is formed from another noun, only prefixation takes place, and when a noun is formed from a verb, both prefixation and suffixation occur.

Ubulawuli is formed from the verb *-lawula* 'to manage/administer' by prefixing class 14 *ubu-* and changing the terminal vowel *-a* into *-i*: *lawula* > *ubu* + *lawula* > *ubulawuli*. *Ubuphathi* is formed from the verb *-phatha* 'administer/manage/control' by prefixing class 14 *ubu-* and changing the terminal vowel *-a* into *-i*: *phatha* > *ubu* + *phatha* > *ubuphathi*. The translator might have utilised the two nouns *ubulawuli* and *ubuphathi* to draw a distinction between 'administration' and 'management' as these are usually treated as synonyms in Xhosa. The derived nouns are rather 'peculiar' to the language. The nouns which are commonly derived from the two verbs are *ulawulo* and *impatho* respectively.

Paraphrasing

Since paraphrasing involves more than one word, sentence alignment is used instead of parallel concordancing. Mtintsilana and Morris (1988: 69) see this strategy as a productive way of extending indigenous vocabularies although the disadvantage is that it consists of more than one word.

The paraphrases given by the translators as shown in Table 10 are acceptable. For instance, for 'marginalised' he used three paraphrases: *zingasiwe* so (those that are not esteemed),

Table 7: Borrowed synonyms

<p>programme > inkqubo/iprogramu</p> <p>DACST (1998) programme > inkqubo <b class=normal1> Line: 957 The publication [[programme]] was diversified, including publication ... <b class=normal2> Line: 1336 [[Inkqubo]] yopapasho yandiswa, kuqukwa nopapasho ...</p> <p>PanSALB (2001/2002) programme > iprogramu b class=normal1> Line: 468 ... by ensuring that films and television [[programmes]] are ... <b class=normal2> Line: 527 ... dla ngokuqinisekisa ukuba [[iprogramu]] zikamabonakude zi</p>
<p>technology > ubuchwepheshe/itekhnoloji</p> <p>DACST (1998) technology > ubuchwepheshe <b class=normal1> Line: 234 The Department of Arts, Culture, Science and [[Technology]] is engaged <b class=normal2> Line: 268 ISebe lezoBugcisa, iNkcubeko, iNzululwazi no[[Buchwepheshe]]</p> <p>PanSALB (2001/2002) technology > itekhnoloji b class=normal1> Line: 808 ... the Special Interest Group for Language and Speech [[Technology]] of the <b class=normal2> Line: N/A ngumbizi-ntlanganiso weqela elinomdla kwimiba ye[[Tekhnoloji]] yoLwimi</p>
<p>telephone > umnxeba/ifowuni</p> <p>PanSALB (2001/2002) telephone > umnxeba <b class=normal1> Line: 789 ...total role in the implementation of the [[Telephone]] Interpreting Service of <b class=normal2> Line: N/A indima ebonakalayo ekukuzalisekiseni inkonzo yotoliko ngo[[mnxeba]] ...</p> <p>SAMSA Manual (2003) telephone > ifowuni b class=normal1> Line: 279 ... 0028 Tel: +27 (0)12 342 3049 <b class=normal2> Line: 280 ... 0028 Ifowuni: +27 (0)12 342 3049</p>

zingananzwanga/zisakunganzwa (those that are not paid attention to) and *zingahoywanga* (those that are neglected). These paraphrases convey more or less the same meaning, as they all refer to something that is not considered important.

For example, 'White Paper', which is paraphrased as *umqulu ongomthetho osacetywayo* (a volume on a statute that is still in draft), *iphepha lengcacisonkqubo* (a paper that clarifies the process/policy) and *iphepha lengxelo-ngqangi likarhulumente* (a paper on the first report of government), is defined as 'an official government report which sets out the government policy on a matter that is or will come before parliament' (*Collins Concise English Dictionary*, 2004). 'Workshop', which is defined as 'a group of people engaged in a study or work on a creative project or subject' (*Collins Concise English Dictionary*, 2004) is paraphrased as *iindibano zokufundisana* (gatherings for teaching one another) or *iindibano zengcaciso* (gatherings for clarification). The translator of the PanSALB Annual Report (2001/2002) did not confine himself to one paraphrase.

Table 8: Terms formed by compounding**PanSALB (2001/2002)****1. accounting > ubalo-mali** (counting of money)

<b class=normal1> Line: 121

. The following are the principal **[[accounting]]** policies used by the Pan South

<b class=normal2> Line: N/A

Le ilandelayo yimigaqo-nkqubo yo**[[balo-mali]]** esetyenziswe yiBhodi yeeLwimi

2. accounting officer > igosa lobalo-mali

<b class=normal1> Line: 16

.. The **[[Accounting Officer]]** is responsible to keep full and proper records ...

<b class=normal2> Line: N/A

[[IGosa loBalo-mali]] linoxanduva lokugcina iirekodi ezipheleleyo ...

ICD Manual (2003)**3. accounting > ucalulo-mali** 'analyzing money'

<b class=normal1> Line: 437

... (d) **[[Accounting]]** records; (...

<b class=normal2> Line: N/A

(d) lingxelo ezibhaliweyo zo**[[calulo-mali]]**;

4. accounting officer > igosa localulo-mali

<b class=normal1> Line: 225

The ICD has its own **[[Accounting Officer]]** (the Executive Director) and ...

<b class=normal2> Line: N/A

(b) I-ICD ine**[[Gosa loCalulo-mali]]** layo (uMphathi Olawulayo) kwaye

5. budget > ulwabiwo-mali 'distributing money'

<b class=normal1> Line: 538

. (e) Records in respect of the **[[budget]]** of the office of the Executive Director;

<b class=normal2> Line: N/A

(e) lingxelo ezibhaliweyo ezimalunga no**[[lwabiwo-mali]]** lwe-ofisi yoMphathi

Short Guide (1998)**6. budgeting > uhlahlolwabiwomali** 'cutting into portions and distribution of money'

<b class=normal1> Line: 528

[[Budgeting]], accounting and reporting: Generally accepted accounting ...

<b class=normal2> Line: 1006

[[Uhlahlolwabiwomali]], ucwangciso nokwenza ingxelo: Inkqubo eqhelekileyo

7. investment > utyalomali 'planting of money'

<b class=normal1> Line: 461

Intergovernmental transfers. Private sector **[[investment]]** ...

<b class=normal2> Line: 872

Ukudluliselwa phakathi koorhulumente. **[[Utyalomali]]** lwecandelo labucala.

PanSALB (2001/2002)**8. policy > umgaqo-nkqubo PanSALB (2001/2002)**

<b class=normal1> Line: 364

... of learning and teaching. The **[[policy]]** requires all schools, wherever possible

<b class=normal2> Line: 414

... zokufundisa nokufunda ulwimi lwesibini. **[[Umgaqo-nkqubo]]** ukhuthaza

9. strategy > isicwangciso-qhinga

<b class=normal1> Line: 708

To develop and implement a **[[strategy]]** on incentives to reward the practising

<b class=normal2> Line: 784

Ukuqulunqa nokusungula **[[isicwangciso-qhinga]]** sokuwonga abo

10. workshop > uqeqesho-msebenzi

<b class=normal1> Line: 433

members of the media who attended the **[[workshop]]** recommended that

<b class=normal2> Line: 488

nawawekho kolu **[[qeqesho-msebenzi]]** benza isindululo sokuba kubanjwe

Table 9: Terms formed by derivation**Short Guide (1998)****1. administration > ubulawuli**

<b class=normal1> Line: 390

The **[[administration]]** has the task of delivering quality services to residents.

<b class=normal2> Line: N/A

Ubulawuli bunexanduva lokunikela ngeenkonzo zodidi kubahlali.

2. management > ubuphathi

b class=normal1> Line: 390

Administrations include the **[[management]]** and

<b class=normal2> Line: N/A

Ulawulo lukamasipala luquka **[[ubuphathi]]** ne

Table 10: Terms formed by paraphrasing

Source text	Target text
<p>1. infrastructure Short Guide (1998)</p> <p>...but involve large-scale capital investment in infrastructure...</p> <p>infrastructure DACST 1998)</p> <p>The institute received infrastructure funding for a drama diploma course, ...</p>	<p><i>izakhiwo zesiseko</i> 'structures of the base'</p> <p>...<i>kodwa ziquka utyalomali oluphezulu ngokuphathelele kwizakhiwo zesiseko</i> ...</p> <p><i>izibonelelo zokuqala ukusebenza</i> 'facilities to start working'</p> <p><i>Eli ziko lafumana uncedo lwemali lokuseka izibonelelo zokuqala ukusebenza ukuze kubekho imfundo yeziqiniselelo zedrama, ...</i></p>
<p>2. marginalised PanSALB (2001/2002)</p> <p>(a) ... vehicle for developing the previously marginalised languages.</p> <p>(b) Conference to discuss the development of marginalised languages: 19 and 20 March 2002.</p> <p>(c) The theme of the conference was 'the position of previously marginalised languages in South Africa, then and now.'</p> <p>marginalised DACST (1998)</p> <p>... the development of historically marginalised languages enjoys a high priority</p>	<p>(a) <i>zingasiwe so</i> 'those that are not esteemed', lit. 'those that are not looked at'</p> <p>(b) <i>zingananzwanga/zisakungananzwa</i> 'those that are not paid attention to'</p> <p>(c) <i>zingahoywanga</i> 'those that are neglected'</p> <p>(a) ... <i>njengeqonga elibalulekileyo lokuphuhlisa iilwimi ezazifudula zingasiwe so.</i></p> <p>(b) <i>Inkomfa engokuphuhliswa kweelwimi ezazifudula zingananzwanga neyayibanjwe ngomhla we-19 nowe-20 kweyoKwindla ku2002.</i></p> <p>(c) <i>Umxholo wenkomfa yaba 'kukukhangela iindawo ezazifudula zikuyo iilwimi ezazingahoywanga zoMzantsi Afrika ngoko nangoku.'</i> <i>ezi Jongelwa phantsi</i> 'those that are undermined'</p> <p>... <i>inkqubela-phambili yeelwimi ezi Jongelwa phantsi ifumana elona lungelo liphambili.</i></p>
<p>3. privatisation Short Guide (1998)</p> <p>Privatisation involves selling municipal assets ...</p>	<p><i>ukunikela kwicandelo labucala</i> 'to hand over to private sector'</p> <p>Ukunikela kwicandelo labucala kuquka ukuthengiswa kwezinto zikamasipala...</p>
<p>4. stakeholders PanSALB (2001/2002)</p>	<p>(a) <i>ababandakanyekayo</i> 'those involved'</p> <p>(b) <i>abachaphazelekayo</i> 'those affected'</p>

Table 10: (continued)

(a) ... were also given to stakeholders seeking advice on language planning development.	(a) ...nazo zanikezelwa kwababandakanyekayo kucelwa iingcebiso ngokuyilwa kokuphuhliswa kolwimi.
(b) the Board to gather input from various stakeholders on how these languages should be developed ...	(b) ...iBhodi ekuqokeleleni izimvo ezivela kwabachaphazelekayo malunga nendlela emakuphuhliswe ngayo iilwimi ...
5. symposium DACST (1998)	<i>indibano yengxoxo</i> 'gathering for discussion' <i>Ngomhla wama-30-31 kuCanzibe1997</i>
On 30–31 May 1997 a symposium with the theme 'The French Presence in South Africa' was held in the auditorium of the museum	indibano yengxoxo emxholo othi 'The French Presence in South Africa' uye wabanjelwa kwelinye lamagumbi alo Vimba.
6. white paper DACST (1998)	<i>Umqulu ongomthetho osacetywayo</i> 'A volume on a statute that is still in draft'
The White Paper on National Water policy for SA	Umqulu ongomthetho osacetywayo
White Paper Short Guide (1998)	<i>wenkqubo yamanzi eSizwe soMzantsi Afrika</i> <i>Iphepha leNgcacisonkqubo</i> 'A paper that clarifies the process/policy'
This section of the White Paper deals with how to restructure the system ...	Eli candelo lePhepha leNgcacisonkqubo
White Paper DACST (1998)	<i>lijongene nendlela yolwakhiwo ngokutsha lwenkqubo ...</i> <i>Iphepha lengxelo-ngqangi likarhulumente</i> 'A paper on the first report of government'
The White Paper on Science and Technology published in 1996, ...	Iphepha leNgxelo-ngqangi likarhulumente elixubusha imiba yezeNzululwazi nobuChwepheshe elishicilelwe ngowe-1996, ...
7. workshops DACST (1998)	<i>iindibano zokufundisana</i> 'gatherings for teaching one another'
The institute received infrastructure funding for a drama diploma course, which included workshops and training of trainers.	<i>Eli ziko lafumana uncedo lwemali lokuseka izibonelelo zokuqala ukusebenza ukuze kubekho imfundo yeziqiniseko zedrama, apho kubakho neendibano zokufundisana, noqeqesho lwabaqeqeshi.</i>
workshops DACST (1998)	<i>iindibano zengcaciso</i> 'gatherings for clarification'
A participative process of discussions and workshops involving the staff of both offices is underway.	<i>Inkqubo engeengxoxo kwakunye neendibano ezingengcaciso ezibandakanya abasebenzi bezi ofisi zimbini iyaqhubeka.</i>

The main advantage of paraphrasing as a strategy is that it achieves a high level of accuracy, although it makes the target text much longer than the source text. Sager (1992: 109) mentions that if a translator cannot find an equivalent, she should paraphrase in order to maintain the informative content of the source language message.

Conclusion

Firstly, in this article the impact of corpus-based research on translation studies as well as case studies on using parallel corpora in translation research were outlined. It was illustrated how Paraconc is used in analysing parallel texts. The successful analysis of parallel texts depends on alignment. Alignment creates links between the source text and the target text. Parallel concordancing enables the researcher to search for a word or phrase and the results are displayed in two windows along with the context. Secondly, it was observed that the translators created terms by using a loan word, using a borrowed synonym, compounding, derivation and paraphrasing. Four types of loan words, i.e. an indigenised loan word, an indigenised loan word preceded by explanation, a pure loan word and a pure loan word preceded by an explanation were identified. Loan words were used mainly for modern computer terms as well as for financial and other

technical terms. Using a loan word and compounding seem to be successful ways of creating technical terms in Xhosa.

References

- Baker M.** 1993. Corpus linguistics and translation studies: Implications and applications. In: Baker, M, Francis G & Tognini-Bonelli E (eds). *Text and Technology: In Honour of John Sinclair*. Amsterdam: John Benjamins, pp 233–250.
- Baker M.** 1995. Corpora in translation studies: An overview and some suggestions for future research. *Target* 7(2): 223–243.
- Barlow M.** 1995. *Paraconc: A concordancer for parallel texts* (Reprinted from *Computers and Texts*.) Houston: Rice University.
- Barlow M.** 2003. *Paraconc: A concordancer for parallel texts (Draft3/03)*. Houston: Rice University.
- Biber D, Conrad S & Reppen R.** 1998. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Bowker L & Pearson J.** 2002. *Working with Specialized Language: A Practical Guide to Using Corpora*. London: Routledge. *Collins Concise Dictionary* 2004. Glasgow: Harper Collins.
- Cluver AD de V.** 1989. *A Manual of Terminography*. Pretoria: Human Sciences Research Council.
- Collins Concise English Dictionary.** 2004. Glasgow: Harper Collins.
- Corporate Communications Department.** 2000. *Eskom glossary of energy terms: English-isiZulu-isiXhosa*. Johannesburg: Eskom.
- Department of Constitutional Development.** 1998. A Short Guide White Paper on Local Government. Pretoria: Arcadia.
- Fischer A, Weiss E, Mdala E & Tshabe S.** 1992. *English-Xhosa Dictionary*. Cape Town: Oxford University Press.
- Gauton R & De Schryver G-M.** 2004. Translating technical texts into Zulu with the aid of multilingual and/or parallel corpora. *Language Matters* 35(1): 133–147.
- Kenny D.** 1997. Abnormal Translations: a German-English Parallel corpus for Investigating Normalisation in Translation. In Lewandowska-Tomaszczyk B & Melia PJ (eds) *Practical Applications in Language Corpora. PALC '97 Proceedings*. Łódź: Łódź University press, pp 387–392.
- King P.** 1997. Parallel corpora for translator training. In: Lewandowska-Tomaszczyk B & Melia PJ (eds). *Practical Applications in Language Corpora. PALC' 97 Proceedings*. Łódź: Łódź University Press, pp 393–402.
- Kraif O.** 2002. Translation alignment and lexical correspondences. In: Altenberg B & Granger S (eds). *Lexis in Contrast: Corpus-based Approaches*. Amsterdam: John Benjamins, pp 271–288.
- Laviosa-Braithwaite S.** 1996. The English Comparable Corpus (ECC): A resource and methodology for the empirical study of translation. Unpublished PhD thesis. Manchester: Department of Language Engineering, UMIS.
- Madiba RM.** 2004. Parallel corpora as tools for developing the indigenous languages of South Africa, with special reference to Venda. *Language Matters* 35(1): 133–147.
- Mtintsilana PN & Morris R.** 1988. Terminography in African languages in South Africa. *South African Journal of African Languages* 8(4): 109–113.
- Pahl H, Pienaar AM & Ndungane TA.** 1989. *The Greater Dictionary of Xhosa* (Volume 3). Alice: University of Fort Hare.
- Pan South African Language Board.** 2002. Annual Report 2001/2002. Pretoria: Arcadia.
- Pearson J.** 2003. Using parallel texts in the translator training environment. In: Zanettin F, Bernardini S & Stewart D. *Corpora in Translator Education*. Manchester: St Jerome, pp 15–24.

- Republic of South Africa.** 1996. *The Constitution of the Republic of South Africa*. Pretoria: Government Printers.
- Sager JC.** 1992. The translator as terminologist. In: Dollerup C & Loddegaard A. *Teaching Translation and Interpreting: Training, Talent and Experience*. Amsterdam: John Benjamins, pp 107–122.
- Satyo SC.** 1981. *Uphengululo lwesiXhosa*. Pretoria: De Jager-Haum Publishers.
- Sinclair J.** 1995. Corpus typology: A framework for classification. In: Mechers G & Warren B (eds). *Studies in Anglistics. Acta Universitatis Stockholmienses*. Stockholm: Almqvist & Wicksell, pp 17–33.
- Trew R.** 1994. The development of training models for African language translators and interpreters. In: Kruger A (ed). *New Perspectives on Teaching Translators and Interpreters in South Africa*. Pretoria: University of South Africa, pp 73–102.
- Wallmach K & Kruger A.** 1999. 'Putting a sock on it': A contrastive analysis of problem-solving translation strategies between African and European languages. *South African Journal of African Languages* **19**(4): 276–289.

Addendum 1: List of indigenised loan words**1. draft > idrafti PanSALB (2001/2002)**

b class=normal1> Line: 599

Final **[[Draft]]** South African Languages Bill 17 May ...

<b class=normal2> Line: 668

[[IDrafti]] yokuGqibela yoMthetho ongeelwimi eMzantsi Afrika 17 kuCanzibe

2. electronic > elektronika PanSALB 2001/2002

<b class=normal1> Line: 258

... **[[electronic]]** material without problems during the process of corpora building...

<b class=normal2> Line: N/A

.. gocagoca iincwadi noovimba be**[[elektronika]]** ngaphandle

3. fax > ifeksi SAMSA Manual (2003)

b class=normal1> Line: 279

[[Fax]]: +27 (0)12 342 3160 SAMSA Durban

<b class=normal2> Line: 280

. **[[Ifeksi]]:** +27 (0)12 342 3160 SAMSA eThekwini

4. internet > intanethi SAMSA Manual (2003)

b class=normal1> Line: 162

... * Hardware * Software ***[[Internet]]** *

<b class=normal2> Line: 162

* (ihadiwe) * linkqubo eziqulethwe yikhompyutha (isoftiwe) ***[[I-intanethi]]**.

5. panel > iphaneli PanSALB (2001/2002)

b class=normal1> Line: 803

... and Technology appointed an advisory **[[panel]]** to oversee the development ...

<b class=normal2> Line: 887

... noBuchwepeshe watyumba **[[iphaneli]]** yabacebisi ukuqinisekisa ukumiselwa ...

6. poster/s > i(i)powusta SAMSA Manual (2003)

b class=normal1> Line: 251

... * Advisory and promotional pamphlets/brochures/**[[posters]]***

<b class=normal2> Line: N/A

*Amaphetshana/iibhrowutsha/ **[[iipowusta]]** zengcebiso nezokwazisa*

7. project > iprojekti PanSALB (2001/2002)

b class=normal1> Line: 223

This **[[project]]** is the brainchild of PanSALB and the Department ...

<b class=normal2> Line: 257

. Le **[[projekti]]** icingwe yiPanSALB neSebe

8. receipt > irisithi

b class=normal1> Line: 333

..A **[[receipt]]** will be issued for cash payments ...

<b class=normal2> Line: N/A

Kuya kukhutshwa **[[irisithi]]** apho kuhlawulwe ngemali esesandleni

9. record > irekodi PanSALB 2001/2002

<b class=normal1> Line: 16

.. The Accounting Officer is responsible to keep full and proper **[[records]]** ...

<b class=normal2> Line: N/A

IGosa loBalo-mali linoxanduva lokucina **[[iirekodi]]** ezipheleleyo ...

10. technology > iteknoloji PanSALB (2001/2002)

b class=normal1> Line: 808

... the Special Interest Group for Language and Speech **[[Technology]]**

<b class=normal2> Line: N/A

gumbizi-ntlanganiso weqela elinomdla kwimiba ye**[[Teknoloji]]** yoLwimi neNtetho

11. telephone > ifowuni SAMSA Manual (2003)

b class=normal1> Line: 279

... 0028 Tel: +27 (0)12 342 3049

<b class=normal2> Line: 280

... 0028 Ifowuni: +27 (0)12 342 3049