

---

# Towards Strategies for Translating Terminology into all South African Languages: A Corpus-based Approach

Rachéle GAUTON<sup>°</sup>, Elsabé TALJARD<sup>‡</sup> & Gilles-Maurice DE SCHRYVER<sup>#</sup>  
*Department of African Languages, University of Pretoria, SA<sup>°‡#</sup> &  
Department of African Languages and Cultures, Ghent University, Belgium<sup>#</sup>*

---

## 1. Introduction

The single biggest problem that translators who translate from a language such as English into the African languages have to contend with is the lack of terminology in the African languages in the majority of specialist subject fields. The relevance of terminology theory and practice for translators therefore becomes clear when the translator is faced with a situation where he/she can no longer rely on existing knowledge and/or dictionaries, and has to conduct research beyond the dictionary.

There is a clear difference between translating into an international language such as English and translating into so-called ‘minor languages’ or ‘languages of limited diffusion’ (LLDs) such as the African languages. This difference also holds regarding the translation of terminology. Cluver (1989: 254) points out that since the terminographer working on a developing language actually participates in the elaboration / development of the terminology, he/she needs a deeper understanding of the *word-formation processes* than his/her counterpart who works on a so-called ‘developed language’.

In this paper, a preliminary study is undertaken, comparing and analysing the various translation strategies utilised by African-language translators in the finding of suitable translation equivalents for English terms foreign to the African languages. To this end, a multilingual corpus of ten parallel texts in all eleven of the official South African languages has been studied. These parallel texts have been culled from the Internet, and a full report on the building of this multilingual corpus can be found in De Schryver (2002). The combined size for all eleven parallel corpora is 348,467 running words, or thus nearly 32,000 words on average per language.

## 2. Methodology followed

The first step in this pilot study is to extract the relevant terminology and to compare the English terms with their translation equivalents in the nine official African languages, *viz.* isiNdebele, siSwati, isiXhosa, isiZulu, Xitsonga, Setswana, Tshivenda, Sepedi and Sesotho, as well as with Afrikaans. For the purposes of this study, we assume that the English texts are the source texts, as all of the websites from which the parallel texts were downloaded, have been written in English, with only small selected sections of the sites in question being provided in the other official languages. Furthermore, it is standard practice in South Africa when undertaking a translation

project involving all nine of the official African languages, to provide the source text in English, as this is in the majority of cases the only language that all of the translators have in common. This is especially the case when the subject matter of the text in question is of a technical nature, as the African languages do not as a rule possess the requisite terminology. On the basis of this evidence it is therefore highly unlikely that any of the African languages (or for that matter Afrikaans) would have served as the source for the texts culled from the Internet on which this study is based.

In extracting the terminology from this corpus of parallel texts, the methodology illustrated by Taljard & De Schryver (2002) is followed. These researchers have shown how African-language terminology can successfully be extracted semi-automatically from untagged and unmarked running text (texts culled from the Internet are, when saved as text files, an example of this) by means of basic corpus query software like *WordSmith Tools*. The key procedure for identifying terminology in each of the parallel corpora is to compare the frequency of every distinct word-type in each parallel corpus, with the frequency of the same word-type in respective reference corpora – the reference corpora obviously being the bigger of the two in each case. Items displaying a great (positive) disparity in frequency are identified as terminology, since the disparity would imply that those specific items occur with unusual (high) frequency in the smaller corpus. The terminology retrieved in this way across the parallel corpora compares very well (see also Uzar & Walinski 2000). For the purposes of this study, the eleven general corpora compiled in the Department of African Languages at the University of Pretoria have been used as reference corpora (for more details, cf. Prinsloo & De Schryver 2002: 256). Sizes of these corpora are typically from a few million up to more than 10 million running words each.

The next step is then to identify the various translation strategies utilised by the different translators in finding suitable translation equivalents for the English terms identified.

### **3. Preliminary results**

On studying the various outputs from the keyword searches done using the two sets of eleven corpora, i.e. the eleven parallel corpora versus the eleven general corpora, the following is readily apparent:

- Although the number of keywords thrown up semi-automatically differs from language to language, there is a good correlation across the parallel corpora between the terms obtained in this manner.
- Even at a casual glance, the following strategies utilised in the translation of source text (ST) terminology are immediately obvious:
  - Translation by means of loanwords in which the English spelling has been retained. Such words have not been transliterated, i.e. nativised in the sense that their phonology has been adapted to reflect the phonological system of the borrowing language.

- Term formation through transliteration. New scientific and technical terms are formed via a process of transliteration by adapting the phonological structure of the loanword to the sound system of the borrowing language.

The occurrence of these two translation strategies in the various languages is summarised in Table 1.

**Table 1:** Keywords and the translation strategies pertaining to loanwords in eleven parallel corpora

Language	Keywords	Loanwords with English spelling		Transliterations	
		#	#	%	#
isiNdebele	583	14	2	37	6
siSwati	427	18	4	16	4
isiXhosa	580	27	5	32	6
isiZulu	619	71	11	30	5
English ST	443				
Afrikaans	426	10	2	17	4
Xitsonga	402	37	9	56	14
Setswana	436	26	6	32	7
Tshivenda	394	43	11	55	14
Sepedi	371	18	5	32	9
Sesotho	320	10	3	13	4

The most important findings regarding these two translation strategies are:

- Whereas isiZulu seems to make use of non-nativised loanwords to a larger extent than transliterations, and whereas siSwati uses these two strategies in equal measure; in all the other languages, i.e. isiNdebele, isiXhosa, the Sotho languages (Setswana, Sepedi and Sesotho), Tshivenda, Xitsonga and Afrikaans, transliterations seem to be used to a greater extent than non-nativised English loanwords as preferred translation strategy for technical terms.
- Many of the non-nativised loanwords under discussion here, are in fact English abbreviations such as SAQA (South African Qualifications Authority), NSB (National Standards Body), RPL (Recognition of Prior Learning), etc. that have been taken over as such into the borrowing language. In Sepedi and Sesotho for example, 78% and 70% respectively of the non-nativised loanwords are English abbreviations that have not been translated into the language concerned, but taken over as is.
- In Afrikaans, translation equivalents are given for English abbreviations such as Eng. SAQA : Afr. SAKO (*Suid-Afrikaanse Kwalifikasie Owerheid*), Eng. NSB : Afr. NSL (*Nasionale Standaardeliggam*), etc., with the noted exception of the abbreviation ANC (African National Congress).

- A similar situation is found in isiNdebele, where translation equivalents are provided for abbreviations such as: Eng. NSB : Ndeb. iHTB (*iHlangano yesiTjhaba yamaBanga*); Eng. SAQA : Ndeb. iPSAF (*UbuPhathimandla beSewula Afrika*), etc.

#### 4. An illustrative example: comparing translation strategies utilised in isiZulu and Sepedi

As was stated at the outset, this paper is intended as a preliminary investigation into strategies utilised in the translation of terminology into *all* South African languages. As this is an ambitious and wide-ranging project, and as time is limited in a forum such as this, two languages, viz. isiZulu and Sepedi, are used as an illustrative example of this process. In Table 2, a representative sample of 20 SL terms are selected from our large database currently under construction, and this is followed by a comparative analysis of the strategies used in the translation of terminology into these languages.

**Table 2:** Comparative analysis of 20 SL items translated into isiZulu and Sepedi

SL term	isiZulu translation equivalent	Sepedi translation equivalent
accreditation	<b>PAU:</b> <i>ukunikezwa amandla / igunya</i> ; <b>ukugunyaza BT:</b> to be given the power / authority, security; to authorise.	<b>MGW:</b> <i>netefatšo</i> <b>BT:</b> verification; <b>MGW:</b> <i>tumelelo</i> <b>BT:</b> permission, approval.
agenda	<b>PAU:</b> <i>uhlelo / uhlu lokuzoxoxwa ngakho</i> <b>BT:</b> arrangement, list of things (issues) that will be talked about / discussed.	<b>MGW:</b> <i>lenaneo</i> <b>BT:</b> list, programme.
apartheid	<b>MGW &amp; MNW:</b> <i>ubandlululo (ngokwebala)</i> <b>BT:</b> discrimination (on the basis of colour), exclusion.	<b>LWT:</b> <i>aparteiti</i> <b>BT:</b> apartheid; <b>MGW:</b> <i>kgethollo</i> <b>BT:</b> separation, segregation - <b>SYN.</b>
assessment criteria	<b>PAR:</b> <i>inqubo yokuvivinyisisa / yokuvivinya</i> <b>BT:</b> criteria (lit. procedure, process) of examining / examining thoroughly; <b>PAR:</b> <i>indlela yokuhlola</i> <b>BT:</b> manner of examining. (All of these paraphrases are rather vague and do not succeed in capturing the exact meaning of the SL term.)	<b>PAR:</b> <i>mokgwa wa tekanyetšo</i> <b>BT:</b> way / manner of estimation; <b>PAR:</b> <i>dinyakwa tša tlhahlobo</i> <b>BT:</b> requirements of examination.
census	<b>LWE:</b> <i>i-census</i> ; <b>MGW:</b> <i>ubalo</i> <b>BT:</b> count (n) - <b>SYN.</b>	<b>MGW:</b> <i>palo</i> <b>BT:</b> count (n.).
definitions	<b>MGW:</b> <i>izincazelo</i> <b>BT:</b> explanations.	<b>MGW:</b> <i>dithlalošo</i> <b>BT:</b> explanations, meanings.
documentation	<b>MGW:</b> <i>izincwadi</i> <b>BT:</b> letters, books; <b>MGW:</b> <i>amabhuku</i> <b>BT:</b> books.	<b>LWT:</b> <i>ditokumente</i> <b>BT:</b> documents.
finance / financial	<b>MGW:</b> <i>izimali / wezimali</i> <b>BT:</b> money / of money.	<b>MGW:</b> <i>(wa / tša) tšhelete</i> <b>BT:</b> (of) money.
gender	<b>RTE:</b> <i>ubulili</i> <b>BT:</b> gender.	<b>RTE:</b> <i>bong</i> <b>BT:</b> gender.
global	<b>PAU:</b> <i>umhlaba wonke jikelele</i> <b>BT:</b> the whole earth, world.	<b>PAU:</b> <i>lefase ka bophara</i> <b>BT:</b> the world at large.

guidelines	<b>COM:</b> <i>imihlahlandlela</i> <b>BT:</b> < - <i>hlahla</i> ‘guide’ + <i>(i)ndlela</i> ‘way, manner’ (Note that the same term is also used to designate ‘framework’.); <b>COM:</b> <i>imikhombandlela</i> <b>BT:</b> < - <i>khomba</i> ‘show’ + <i>(i)ndlela</i> ‘way, manner’.	<b>COM:</b> <i>methalohlahli</i> <b>BT:</b> < <i>methala</i> ‘lines’ + <i>hlahla</i> ‘guide’; <b>COM:</b> <i>ditšhupatsela</i> <b>BT:</b> < <i>šupa</i> ‘show’ + <i>tsela</i> ‘road, way’.
institutions	<b>SSP:</b> <i>izikhungo</i> <b>BT:</b> (lit.) gathering places.	<b>LWT:</b> <i>diinstithušene</i> <b>BT:</b> institutions.
Minister	<b>RTE:</b> <i>ungqongqoshe</i> <b>BT:</b> minister.	<b>CST:</b> <i>tona</i> <b>BT:</b> advisor to the chief / king.
outcome(s)	<b>MGW:</b> <i>imiphumela</i> <b>BT:</b> results; <b>RTE:</b> <i>impumelelo</i> <b>BT:</b> outcome, success.	<b>MGW:</b> <i>dipoelo</i> <b>BT:</b> results.
redress	<b>MGW:</b> <i>ukulungisa</i> <b>BT:</b> to correct, rectify.	<b>MGW:</b> <i>phetolo</i> <b>BT:</b> change, reversal.
regulation(s)	<b>COM:</b> <i>imithethonkambiso</i> <b>BT:</b> < <i>imithetho</i> ‘laws, rules’ + <i>(i)nkambiso</i> ‘custom’.	<b>MGW:</b> <i>melawana</i> <b>BT:</b> small laws.
research	<b>RTE:</b> <i>ucwaningo</i> <b>BT:</b> research.	<b>LWT:</b> <i>resetšhe</i> <b>BT:</b> research; <b>SSP:</b> <i>nyakišišo</i> <b>BT:</b> investigation - <b>SYN.</b>
South African Qualifications Authority	<b>LWE:</b> <i>i-South African Qualification(s) Authority</i> ; <b>PAR:</b> <i>Isigungu seziPhathimandla sokwengamela iziqu eNingizimu Afrika</i> <b>BT:</b> authorising committee that presides over South Africa's qualifications - <b>SYN.</b>	<b>PAR:</b> <i>Bolaodi bja Mangwalo a Thuto bja Afrika Borwa</i> <b>BT:</b> authority of letters of learning of South Africa.
stakeholder(s)	<b>MNW:</b> <i>abathintekayo</i> <b>BT:</b> those affected.	<b>COM:</b> <i>bakgathatema</i> <b>BT:</b> those who take part.
Standards Generating Body	<b>LWE:</b> <i>iStandards Generating Body</i> ; <b>PAR:</b> <i>uMgwamanda eKhiqiza / eYenza amaZinga</i> <b>BT:</b> assembly, congregation, community that (abundantly) produces / makes standards; <b>RTE</b> & <b>LWE:</b> <i>uMgwamanda iStandards Generating Body</i> - <b>SYN.</b>	<b>PAR:</b> <i>Lekgotla la Tlhamo ya Maemo</i> <b>BT:</b> council of establishment of standards.

Note that in Table 2, the SL terms are listed as proffered by the keyword search, i.e. in derived or inflected form. However, should a terminology list be compiled, these terms will be lemmatised under their canonical forms. Note also that the following codes are used to symbolise the strategies that are, according to Baker (1992: 26-42), often used by professional translators in solving various types of problems of non-equivalence at word-level:

- **MGW:** Translation by a more general word (superordinate).
- **MNW:** Translation by a more neutral or less expressive word.
- **CST:** Translation by cultural substitution.
- Translation using a loan word or loan word plus explanation (sometimes in brackets):
  - **LWE:** Translation by means of loanwords in which the English spelling has been retained. Such words have not been transliterated, i.e. nativised in the sense that their phonology has been adapted to reflect the phonological system of the borrowing language.

- **LWT:** Term formation through transliteration. New scientific and technical terms are formed via a process of transliteration by adapting the phonological structure of the loanword to the sound system of the borrowing language.
- **PAR:** Translation by paraphrase using a related word, i.e. paraphrasing by using a direct / ready equivalent of the SL item in the paraphrase.
- **PAU:** Translation by paraphrase using unrelated words, i.e. paraphrasing by not using a direct / ready equivalent of the SL item in the paraphrase.

In addition to the translation strategies listed above, it is well known that translators working into the African languages are more often than not required to create new terms, and should therefore be completely *au fait* with term creation strategies in their particular language. Regarding term formation in the African languages, Mtintsilana & Morris (1988: 110-112) distinguish between *term-formation processes* internal to the language, and borrowings from other languages. They identify a number of term formation processes in the African languages, of which the following appear in Table 2:

- Semantic transfer: This is the process of attaching new meaning to existing words by modifying their semantic content.
  - **SSP:** In the creation of new terms, the most common form of semantic transfer is *semantic specialisation*, i.e. a word from the general vocabulary acquires an additional, more technical meaning.
- **COM:** Compounding. The term is coined by combining existing words.
- **SYN:** Synonym richness of the vocabulary. Although this is not a method of creating new terms, Mtintsilana & Morris point out that the relative abundance of synonyms in African-language vocabularies offers both advantages and disadvantages from a terminological point of view. E.g., a term may be coined for a foreign concept while a transliteration of the foreign term is also in use.
- Lastly, in some cases in Table 2 above, there is no problem of non-equivalence (at word level) between the source and target languages, as the TL possesses a ready translation equivalent of the SL term in question. Such cases are designated with the code **RTE** (ready translation equivalent).
- The code **BT** in Table 2 above, stands for Back-translation.

The data from Table 2 is quantified in Table 3. (Note that in cases where there are two translation equivalents for a particular keyword, each of these equivalents is counted as a half).

**Table 3:** Quantitative analysis of 20 SL items translated into isiZulu and Sepedi

Translation strategy	isiZulu		Sepedi	
	# terms	% terms	# terms	% terms
Paraphrase:				
• PAR	2	10	3	15
• PAU	3	15	1	5
Term formation strategies:				
• LWE	1.5	7.5	—	—
• LWT	—	—	3	15
• COM	2	10	2	10
• SSP	1	5	0.5	2.5
More general and/or neutral word:				
• MGW	5.5	27.5	8.5	42.5
• MNW	1.5	7.5	—	—
RTE	3.5	17.5	1	5
CST	—	—	1	5
<b>Total</b>	<b>20</b>	<b>100</b>	<b>20</b>	<b>100</b>

The following conclusions can be drawn from Table 3:

- In both isiZulu and Sepedi, translation by a more general and/or neutral word seems to be the preferred strategy, i.e. in a little more than a third of all cases (35%) in the isiZulu sample and approaching half of the cases (42.5%) in the Sepedi sample.
- The next most popular translation strategy in isiZulu would seem to be translation by paraphrase at 25% of the sample.
- This contrasts with Sepedi where term formation is utilised in just over a quarter of the cases (27.5%) as the next most popular translation strategy after translation by a more general word.
- Term formation as translation strategy is found in just less than a quarter of cases in the isiZulu sample (22.5%).
- In Sepedi, translation by paraphrase accounts for another fifth of the sample (20%).
- In only 17.5% of the cases does isiZulu make use of a ready / direct translation equivalent, whilst in Sepedi the remaining 10% of the cases consists of one instance of translation through the use of a ready / direct equivalent, and one instance of translation through cultural substitution.
- The same translation strategy is used in both isiZulu and Sepedi in the translation of the following SL terms: *assessment criteria*, *definitions*, *finance / financial*, *gender*, *global*, *guidelines* and *redress*.
- In a few cases, both isiZulu and Sepedi display synonym richness. This is the case with the SL terms *census*, *South African Qualification(s) Authority* and *Standards Generating Body* in isiZulu, and *apartheid* and *research* in Sepedi.

## 5. Conclusion

In this paper we have shown how electronic machine-readable corpora can be used in determining the strategies used by professional translators in finding translation equivalents for SL terms. This is a wide-ranging project that will require the participation of researchers from all of the South African languages, and which will on completion provide a wealth of data with numerous practical applications. Apart from the obvious benefits of this undertaking for the fields of translation studies and terminology, the results from this project will provide guidelines to especially African-language translators, confronted with the onerous task of finding translation equivalents for SL terms foreign to these languages.

## References

- Baker, M.** 1992. *In other words: a coursebook on translation*. London: Routledge.
- Cluver, A.D. de V.** 1989. *A manual of terminography*. Pretoria: Human Sciences Research Council.
- De Schryver, G.-M.** 2002. Web for/as Corpus: A Perspective for the African Languages. *Nordic Journal of African Studies* 11/2: 266-282.
- Mtintsilana, P.N. and R. Morris.** 1988. Terminography in African languages in South Africa. *South African Journal of African Languages* 8/4: 109-113.
- Prinsloo, D.J. and G.-M. de Schryver.** 2002. Towards an 11 x 11 Array for the Degree of Conjunctivism / Disjunctivism of the South African Languages. *Nordic Journal of African Studies* 11/2: 249-265.
- Taljad, E. and G.-M. de Schryver.** 2002. Semi-Automatic Term Extraction for the African Languages, with special reference to Northern Sotho. *Lexikos* 12: 44-74.
- Uzar, R. and J. Walinski.** 2000. A comparability toolkit: Some practical issues for terminology extraction. In B. Lewandowska-Tomaszczyk and P.J. Melia (eds.). *PALC'99: Practical Applications in Language Corpora. Papers from the International Conference at the University of Lodz, 15-18 April 1999*: 445-457. (Lodz Studies in Language 1.) Frankfurt am Main: Peter Lang.