

Multilingual parallel corpus: An institutional resource for terminology development at the University of South Africa (Unisa)

Koliswa Moropa – Bulelwa Nokele
University of South Africa / South Africa

Abstract –The indigenous African languages of South Africa are not fully developed to provide for specialised terminology and were considered unsuitable for use as languages of tuition and research. This was used as a scapegoat for not utilising these languages in the South African education system. Since 1994, however, terminology development has been one of the key priorities of democratic South Africa. The institutions of Higher Learning have been mandated to develop and intellectualise the indigenous languages for teaching, learning and research. In line with this, this article aims to address the problem of unavailability of scientific or technical terms by illustrating how a multilingual corpus —from which multilingual glossaries as resources for tuition and research— can be compiled. Adopting a qualitative descriptive approach, suitable source texts in English and their translations in various African indigenous languages, namely, IsiZulu, IsiXhosa, IsiNdebele, SiSwati, Tshivenda, and Xitsonga were selected from the University study material for inclusion in the multilingual parallel corpus. *ParaConc*, a software that is suitable to query parallel texts, was used to align and extract terms from the corpus. The study demonstrates how parallel texts can be useful in developing scientific and technical terms. The University of South Africa can become the centre of corpus compilation for the intellectualisation of the official indigenous South African languages, since it is the only university in the country that caters for all these languages.

Keywords –corpus; corpus compilation; terminology development; multilingual parallel corpus; indigenous languages

1. INTRODUCTION

This article provides information on the main topics to be considered when designing a corpus and tries to answer the question of why it is imperative for the University of South Africa (Unisa) to move towards corpus compilation to promote terminology development. Commenting on the important issue of language development and the Unisa Language Policy (2016), Alexander (2003: 18) observes that:

[m]ere language planning cannot bring about the fundamental shifts in consciousness and in behaviour which are necessary to lift the indigenous languages of Africa to a different historical trajectory.



Alexander finds it illogical to believe that it is possible to think of an African renaissance without the development and intellectualisation of the South African indigenous languages. He calls on Higher Education Institutions (HEIs) to participate in facilitating and promoting the development of these languages in such a manner that they can be used in all official functions as formal academic languages in HEIs.

Since 1994, language development has been one of the key priorities of the democratic South Africa. The Constitution of the Republic of South Africa, Act 108 of 1996, brought about changes regarding the status of indigenous African languages, declaring nine of them to be official. Section 6 (1) of the Constitution stipulates that “the official languages of the Republic of South Africa are Sepedi, Sesotho, Setswana, Tshivenda, Xitsonga, Afrikaans, English, SiSwati, IsiNdebele, IsiXhosa, and IsiZulu.” Before that, the education policies of the apartheid regime neglected the indigenous languages, recognising only Afrikaans and English as official languages. Likewise, the colonial bilingual education system marginalised all other indigenous languages and, therefore, the technical/scientific registers of these languages remained underdeveloped (Moropa and Shoba 2017).

1.1. Language Policy for Higher Education

A milestone in the commitment to language development and promotion of multilingualism in institutions of Higher Learning was the *Language Policy for Higher Education* (LPHE; Department of Education 2002), which was revised in Department of Education (2020). The revised version of the policy framework declares that since the propagation of the LPHE in 2002, little progress has been made in exploring and exploiting the potential role of indigenous African languages in facilitating access and success—as well as in the intellectualisation of these languages—in Higher Education. LPHE requires HEIs to develop language policies together with implementation plans.

In 2016, Unisa amended its language policy, and in 2017 it adopted the implementation plan entailing that, as a national university a) it acknowledges that there are eleven official languages in South Africa and ensures that, together with South African Sign Language, they enjoy parity of esteem and equitable treatment (Section 4.1.1), and b) it endeavours to support all the official languages of South Africa (Section 4.1.6).

As an “African university, in the service of humanity, shaping futures,” Unisa commits to building capacity for all official South African languages (Unisa Language Policy 2016, Section 4.1.6). The initiative of corpus compilation as a resource for terminology development is thus a step towards promoting the indigenous African languages as languages of teaching, learning and research at Unisa, and this constitutes one step towards fulfilling this obligation.

1.2. The research problem and aim

Different scholars have argued that one of the reasons hindering the development and intellectualisation of African languages is the lack of terminology (cf. Madiba 2001; Gauton and De Schryver 2004). Different intervention strategies have been suggested to sort this out. Madiba (2001) suggests a pragmatic approach which entails borrowing and indigenisation of terms from source languages. Moropa (2005) recommends the use of the concordancer *ParaConc*¹ for term identification and term creation strategies when dealing with technical texts. Addressing the lack of specialised quadrilateral dictionaries, Mlambo *et al.* (2021) also use *ParaConc* to identify lexical items in English and their equivalents in Xitsonga, SiSwati and IsiNdebele, to compile a quadrilateral dictionary. Their broader aim is to facilitate communication across languages and development of minority languages, which will ultimately promote multilingualism in South Africa. These are some studies that have not only voiced concern about the lack of terminology in the indigenous languages of South Africa, but have also proposed solutions to overcome this problem.

As the largest—and the only—university that serves all language communities in South Africa, Unisa is well positioned to contribute towards the noble mandate of developing all previously disregarded languages to end the negative narrative of lack of terminology. Therefore, the University has embarked on a project involving translating study material from English into all the South African official languages. The translated material is part of the corpus that could be used to extract and/or develop specialised terminology. It is also imperative to note that the translation of these study materials is not only for mere availability in the various indigenous African languages, but for access of information in a language that is understandable to the reader. That is why paraphrasing

¹ <https://paraconc.com/>

is one of the popular translation strategies that has been used. Set within this background, the present article seeks to illustrate how a corpus can assist in the development and intellectualisation of the indigenous languages.

The paper is structured as follows. Section 2 discusses the notion of ‘corpus’ and how corpora are being compiled in South Africa, whereas Section 3 offers a qualitative discussion of the results. Section 4 concludes the article.

2. CORPUS COMPILATION

2.1. Definition of ‘corpus’ and types of corpora

Originally, the term ‘corpus’ (plural ‘corpora’) meant any collection of writings in a processed or unprocessed form by a specific author. According to Baker (1995: 225), with the advancement of corpus linguistics, this definition changed in three important ways:

1. Nowadays, a corpus is primarily a collection of texts held in machine-readable form, which can be analysed (semi)-automatically in a variety of ways.
2. A corpus is no longer restricted to written texts but can also include spoken texts.
3. A corpus may include a number of texts from a variety of sources by many writers and speakers on a multitude of topics.

What is important is that a corpus is assembled for a *particular purpose* (emphasis added) and according to explicit design criteria to ensure that it is *representative* (emphasis added) of the given area or sample of language for which it aims to account for.

In his classification, Sinclair (1995) distinguishes various types of corpora such as reference, monitor, parallel and comparable corpora. A parallel corpus is a “collection of texts, each of which is translated into one or more other languages than the original” (Sinclair 1995: 32). A parallel corpus can be bilingual when it comprises original texts and their translated versions of the same source language. For example, in the South African context it can be English Source Text (original) [ST] – IsiZulu Target Text (translated) [TT]; English [ST] – Tshivenda [TT], or *vice versa*. The parallel corpus is multilingual, as it contains translations into several target languages. For example, tutorial letters in various disciplines are translated from English into other official South African languages to make the texts accessible to all (e.g., English [ST – Setswana – IsiNdebele – Xitsonga [TTs], etc.).

2.2. Corpus-based research for terminology development in South African languages

Corpus-based research for terminology development in various indigenous African languages has been widely conducted in South Africa. Madiba (2004), Gauton and De Schryver (2004), Moropa (2005, 2007), Ndhlovu (2016), and Shoba (2018) have shown how specialised corpora can be used to develop indigenous African languages and retrieve terminological information. Madiba (2004), who analyses the *Special Language Corpora for African Languages* (SPeLCAL), illustrates how parallel corpora can be used as tools for developing the indigenous languages of South Africa. The SPeLCAL 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. Madiba (2004) makes use of *Multiconcord*² to analyse translation equivalents of terms such as 'act', 'legislation', 'rule', 'order', and 'law' in a parallel corpus of English-Tshivenda texts of The Constitution of the Republic of South Africa (1996). His findings revealed that translation equivalents, as well as inconsistencies in the translation, can be identified and that the SpeLCAL corpora could be useful for terminographers and lexicographers.

Gauton and De Schryver (2004) demonstrate how special-purpose multilingual and parallel corpora can be used as a translator's tool in finding suitable equivalent terms when translating technical texts from English into Zulu. In their study they make use of a) the *University of Pretoria Zulu Corpus* (PZC), an electronic corpus compiled at the University of Pretoria which comprises literary texts, religious texts, internet files, and pamphlets in Zulu and a total amount of five million words (first case study), and b) the *University of Pretoria Internet English Corpus* (PIEC), an electronic corpus of 12.4 million English words retrieved from online sources (second case study.)³ In the first case study, multilingual corpora are used to investigate the terminology used in the translation of HIV/AIDS texts. HIV/AIDS terminology was identified in both corpora by resorting to the *KeyWords* function in *WordSmith Tools*.⁴ In the second case study, parallel corpora dealing with labour issues are scrutinised to investigate labour and determine the usefulness of such corpora as a resource for the translation of technical texts into Zulu.

² https://artsweb.cal.bham.ac.uk/pking/multiconc/l_text.htm

³ <https://sadilar.org/index.php/en/>

⁴ https://lexically.net/downloads/version5/HTML/index.html?keywords_start.htm

Moropa (2004, 2005, 2007) also investigates how corpus-based research may contribute to the development of strategies for translating financial and technical texts into IsiXhosa. Moropa's research illustrates the benefits of parallel texts and computer tools, such as *ParaConc*, in the development of terminology and also their usefulness for translators as terminology developers. Proper alignment and good quality translations are also emphasised for translations to be used as resources for terminology development.

Along the same lines, Ndhlovu (2016) resorts to the *English-Ndebele Parallel Corpus* (ENPC; Ndhlovu 2012) to extract bilingual terminology for the creation of an English-Ndebele medical dictionary. The corpus comprises English STs and equivalent Ndebele TTs with multiple translations collected from Zimbabwean non-governmental and governmental organisations.

Likewise, Shoba (2018) explores how parallel corpora can be analysed with the use of the concordancer *ParaConc* to extract bilingual terminology that can be used to create specialised bilingual dictionaries. Shoba (2018) follows a corpus-based approach because it quickly, efficiently, and accurately allows the extraction of bilingual terms in their immediate contexts.

2.3. South African universities and corpus compilation

To illustrate the state of affairs of corpus compilation as an institutional resource in South Africa, we will show two examples of HEIs in the country that have made progress in corpus building: a) the University of Pretoria and b) the University of KwaZulu-Natal. The Department of African Languages of the University of Pretoria has been involved in corpora compilation since the early 1990s. It compiled the *Pretoria Sesotho sa Leboa Corpus* (PSC; Prinsloo 1991), which started with 156,000 running words and comprises millions of words nowadays. Similarly, there is a joint project between the Departments of African Languages of the Universities of Pretoria and Ghent that has generated large corpora for all official South African languages, with sizes averaging several million tokens per language (De Schryver and Daniëlle 2005).

The University of KwaZulu-Natal, in line with its language policy and plan, has established a centre specialising in designing an *IsiZulu National Corpus*⁵ (INC) and an

⁵ <https://inzc.ukzn.ac.za>

IsiZulu term bank as key enablers in the development of Human Language Technologies (HLT). The University of KwaZulu-Natal, through its University Language Development and Planning Office (ULDPO), has undertaken the development of computer programmes and technologies for the study and use of IsiZulu. These include a) INC, which comprises 20 million tokens, b) an IsiZulu term bank for a variety of disciplines, c) an IsiZulu spellchecker software for writing and editing in IsiZulu with interfaces in both IsiZulu and English, and d) an electronic Zulu lexicon. In essence, they focus on developing IsiZulu as a language of teaching and learning at the university (Khumalo *et al.* 2019).

2.4. *Corpus compilation at Unisa*

The *Unisa Corpus* discussed here mainly represents Language for Special Purposes (LSP) since the University has adopted a multilingual approach to the creation of learning, such as the use of glossaries that can assist non-proficient students of English to access specialised subject fields in their preferred language. LSP refers to language that tends to be formal and contains a highly specialised vocabulary. LSP texts are restricted and precise and typically feature, amongst others, an abundance of specialised terms. LSP phrases and terminology are chiefly aimed at serving the communication needs of specialists. LSP language strongly contrasts with Language for General Purposes (LGP), which is of common usage.

In implementing its language policy in 2016, the University of South Africa, through its strategic project *Transformation: Building capacity for South African languages*, started collating Unisa texts that have been translated from English several South African official languages. This collection of texts includes an ST and its TTs and it is the first step in corpus compilation, known as ‘corpus design’. The texts comprise tutorial material. The tutorial texts are grouped according to domains that represent various disciplines: education, economic and management sciences, accounting sciences, human sciences, law, agriculture and environmental studies, science, and engineering and technology, as well as the sub-disciplines within these. As the research progresses, other domains will be determined by the type of texts which the institution generates.

Selecting and collecting suitable texts for inclusion in the corpus, as well as obtaining corpus analysis tools, are crucial steps in corpus-based compilation and

research. A corpus needs to be carefully compiled and the quality of the investigation is directly related to the quality of the data (Granger 1998). A corpus is not simply a collection of texts, but rather the appropriate design of a corpus depends on what it is meant to represent. What is deposited into the corpus determines the outcome. As Leech (1998) aptly notices, paying proper attention to quality and design criteria always takes twice as long as one would think and it may take ten times as much effort. Creating a corpus requires a number of researchers, for example, those responsible for selecting the criteria in identifying the texts to be included, for collecting the texts, for grouping the texts according to the individual domains, medium and time, size, etc. De Schryver and Daniëlle (2005) identify three steps in the process of corpus compilation: a) corpus design, b) text collection, and c) text encoding. The size of a corpus is described in terms of the size of individual texts. For a parallel corpus, it is the size of the ST and TTs and the total number of words in the corpus.

In the present study, the corpus selected contains texts from the disciplines of human sciences and accounting sciences, respectively. Table 1 below shows the names of the individual STs and languages of translation. The author of the STs is identified as the Faculty, and the translator is the Language Unit that manages the translation workflow.

	STs and author/s Faculty of Human Sciences Faculty of Accounting Sciences	TTs Translator: Language Unit
1	Honours in Development Studies Code: DVAALD_2019_TL_301_B)	<i>Izifundo Zentuthuko</i> (IsiZulu)
2	Introductory Financial Accounting Code: (FAC_2022_TL_101_B)	<i>Dithuto tša Tlhabollo</i> (Sepedi) <i>Selelekela ho Akhaonting ya tsa Ditjhelete</i> (Sesotho) Introductory Financial Accounting (Setswana) Introductory Financial Accounting (Sepedi) <i>Intshayelelo ngoCwangciso lwezeMali</i> (IsiXhosa) <i>Isethulo Sokubalisisa Izimali</i> (IsiZulu) <i>Singeniso Kutekuphatfwa Kwetimali</i> (SiSwati) Introductory Financial Accounting (IsiNdebele) Introductory Financial Accounting (Tshivenda) Introductory Financial Accounting (Xitsonga)

Table 1: Translated tutorial letters for DVAALD and FAC

2.5. Corpus tools

The value of corpora as sources of data lies in that the data may be accessed with the use of software tools. Text retrieval programmes, commonly referred to as ‘concordancers’, are the most widely linguistic software tool used. Concordancers allow counting words and sequences of words and sort them in a variety of ways. They also provide information on how words combine with each other in the text. In addition, they can carry out comparisons of entities in two corpora and bring out statistically significant differences. Software tools such as *WordSmith Tools*, *Multiconcord*, *ParaConc*, *Sketch Engine* (Kilgarriff *et al.* 2014), and many others enable researchers to carry out searches, which they could never hope to do manually.

In the present analysis, *ParaConc* has been selected as a software programme designed primarily as a search tool that works with parallel texts. *ParaConc* combined with a suitable set of texts can be used as a full context bilingual dictionary or as linked bilingual discourse of translation equivalences and present the user with a) multiple instances of the search term, and b) a large context for each instance of the search term, thereby allowing a thorough analysis of usage in terms of the equivalences between two languages (Barlow 2008: 12).

To illustrate the application of software tools, we provide some screenshots, one from a Development Studies Tutorial Letter (DVAALD_2019_TL_301_B) translated into IsiZulu (see Figure 1), and another one from Introduction to Financial Accounting (FAC_2022_TL_101_B) translated into SiSwati (see Figure 2). It is worth noticing that the African languages we used for illustration and analysis in this study are IsiNdebele, IsiXhosa, IsiZulu and SiSwati. Since FAC 1501 has more TTs and *ParaConc* takes a maximum of four files at a time —English (ST) and IsiXhosa, IsiNdebele and IsiZulu— SiSwati was loaded separately.

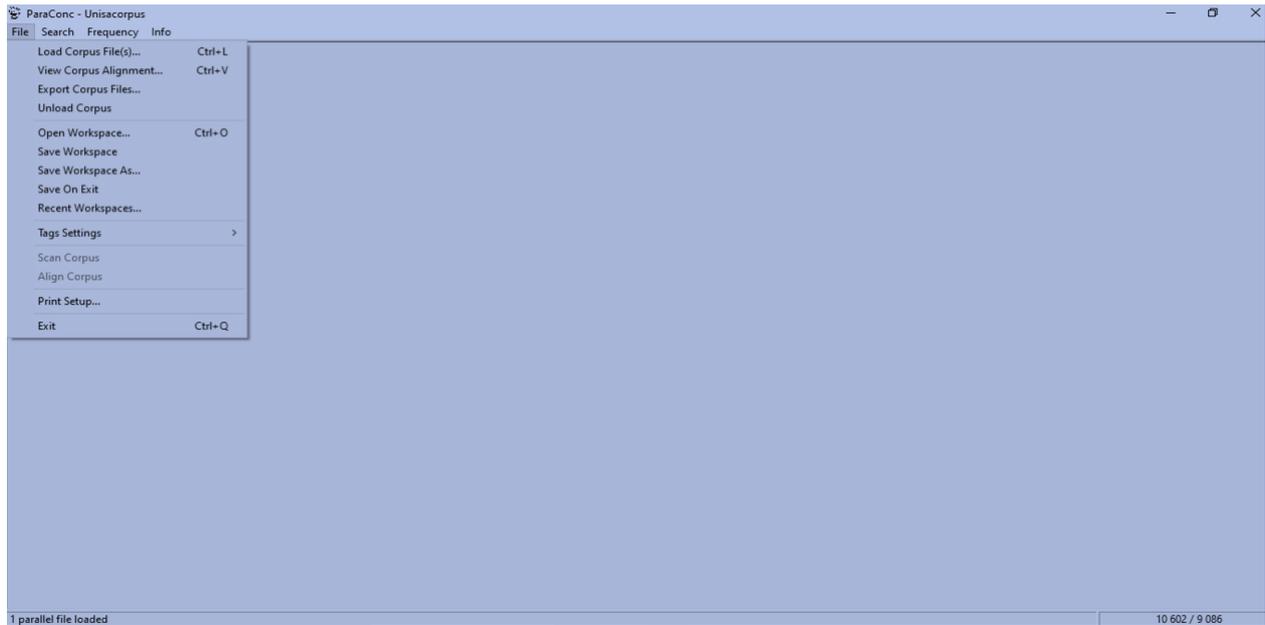
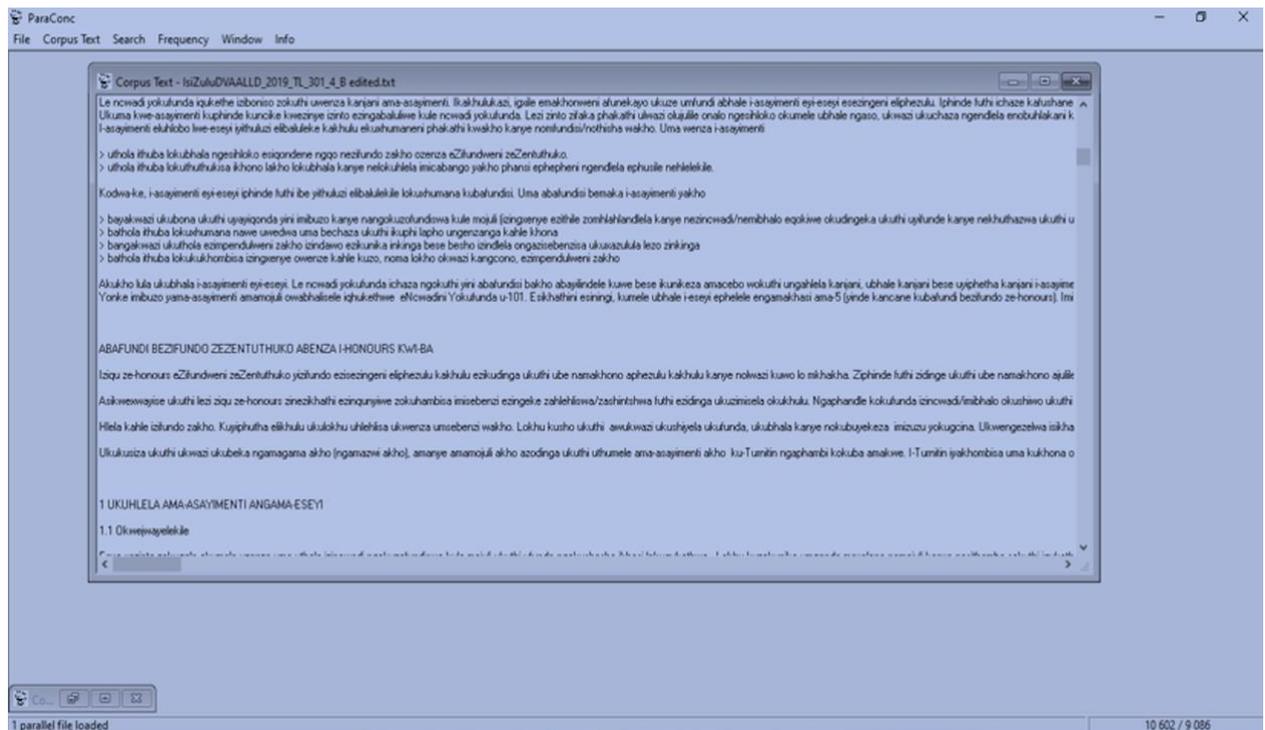
Figure 1: *ParaConc* file menu: DVAALD_2019_TL_301_B

Figure 2: IsiZulu plain text – DVAALD_2019_TL_301_B

The processing of the files may take some time and it is advisable to use the workspace option to minimise the number of times that texts must be uploaded. There is no real limit to the size of the corpus loaded. Even though the corpus files appear to be loaded in the programme ready for searching, in practice, *ParaConc* does not load the whole text, but rather switches chunks of text in and out of memory as one does the searches. This means that the programme should theoretically be able to handle any size of text (see Figure 3

of parallel texts loaded in workspace). Since *ParaConc* does not have a list of the South African indigenous languages, it should be noted that French Canadian is used to label IsiZulu (cf. Figure 3). Finally, Afrikaans, ‘Additional 1’ and ‘Additional 2’ are used as labels in Figure 4. In the FAC1501 corpus files, ‘Afrikaans’ is a label for IsiXhosa, ‘Additional 1’ is a label for IsiNdebele and ‘Additional 2’ is for IsiZulu.

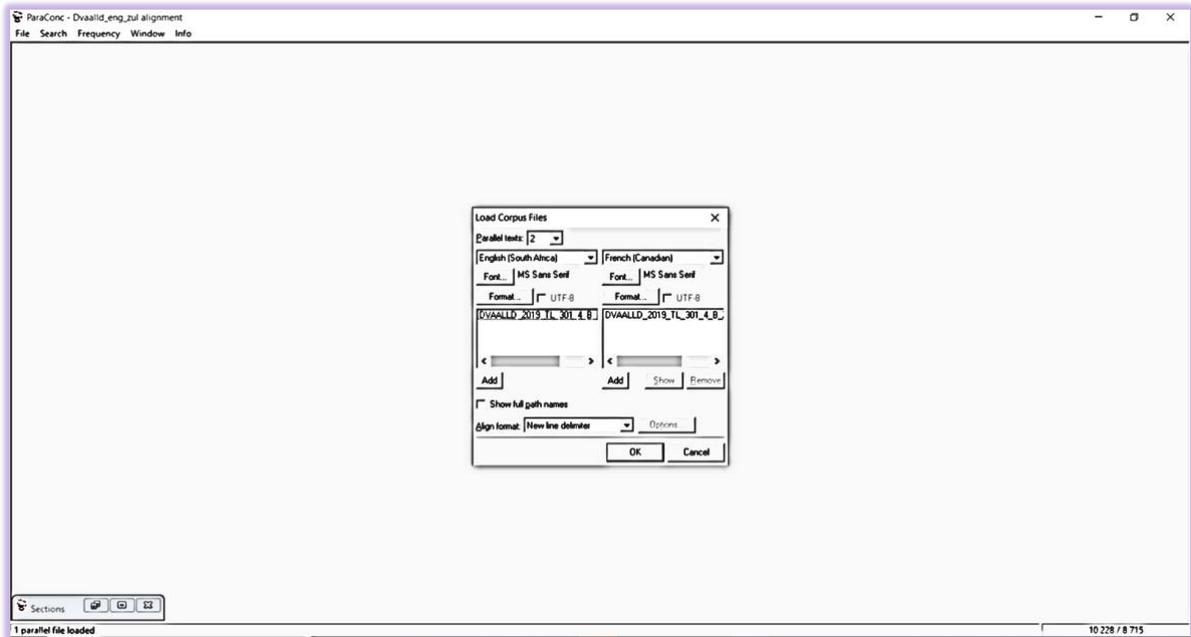


Figure 3: Parallel texts loaded in workspace

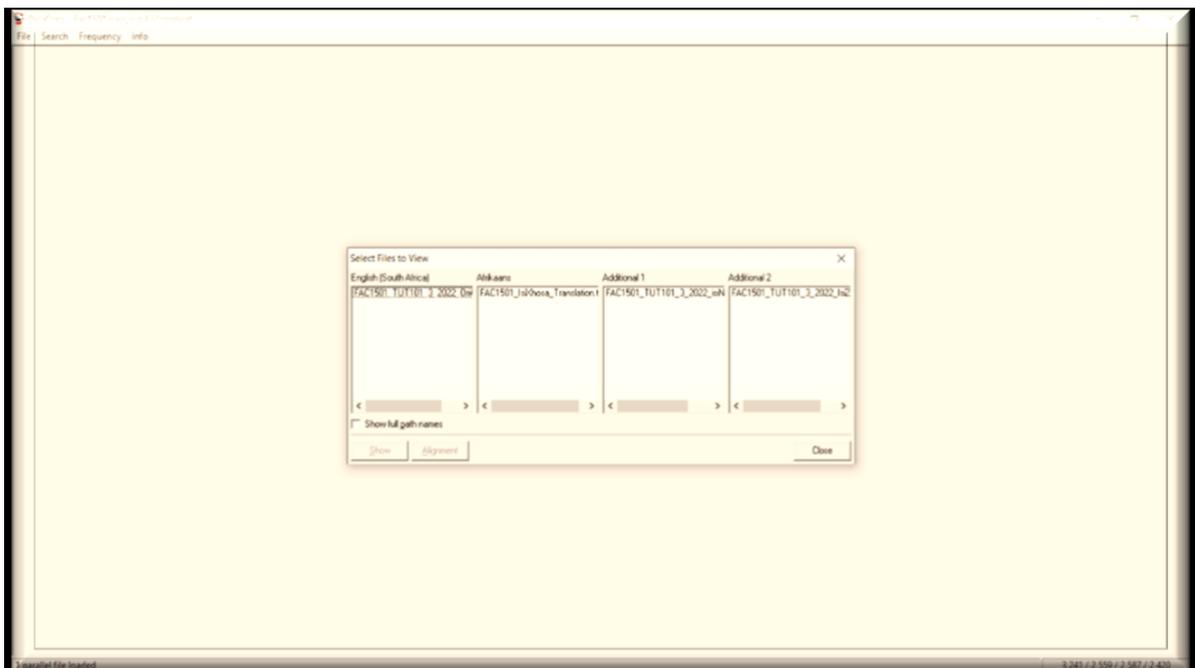


Figure 4: FAC1501 parallel texts loaded in workspace

2.6. Creating a word list

Once the texts have been loaded on *ParaConc*, 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 gives an overall idea of what information can be searched for, and it is very useful for the identification of terminology. The word list may be arranged according to the order of frequency or alphabetically (Figures 5 and 6). The word list in Figure 6 is arranged according to the order of frequency in English, IsiXhosa, IsiNdebele and IsiZulu.

English (South Africa)			French (Canadian)		
Count	Pct	Word	Count	Pct	Word
236	2.3074%	a	12	0.1377%	a
4	0.0391%	abbreviation	11	0.1262%	ababhali
9	0.0880%	ability	4	0.0459%	abacaphuni
7	0.0684%	able	18	0.2065%	abafundi
20	0.1955%	about	12	0.1377%	abafundisi
4	0.0391%	above	4	0.0459%	abahlukene
6	0.0587%	ac	3	0.0344%	abakusiza
5	0.0489%	academic	4	0.0459%	abangaphezu
8	0.0782%	accessed	15	0.1721%	abantu
3	0.0293%	according	7	0.0803%	abanye
6	0.0587%	acknowledge	3	0.0344%	abehlukene
8	0.0782%	act	3	0.0344%	abhalwe
5	0.0489%	action	4	0.0459%	abo
6	0.0587%	add	6	0.0688%	ac
4	0.0391%	advanced	9	0.1033%	acashuniwe
5	0.0489%	advice	8	0.0918%	accessed
5	0.0489%	affairs	6	0.0688%	act
17	0.1662%	africa	5	0.0574%	affairs
5	0.0489%	african	15	0.1721%	africa
9	0.0880%	after	5	0.0574%	african
4	0.0391%	against	4	0.0459%	ake
6	0.0587%	al	30	0.3442%	akho

Figure 5: Alphabetical word list – DVAALLD English – IsiZulu

English (South Africa)		AfiXhosa		Additional 1		Additional 2	
Count	Pct	Count	Pct	Count	Pct	Count	Pct
156	4.8133%	46	1.7976%	32	1.2370%	34	1.4050%
122	3.7643%	28	1.0942%	29	1.1210%	30	1.2397%
96	2.9620%	27	1.0551%	29	1.1210%	30	1.2397%
76	2.3450%	23	0.8988%	27	1.0437%	27	1.1157%
69	2.1290%	22	0.8597%	27	1.0437%	27	1.1157%
64	1.9747%	22	0.8597%	220	8504%	22	0.9091%
59	1.8204%	21	0.8206%	220	8504%	22	0.9091%
53	1.6353%	20	0.7816%	220	8504%	21	0.8678%
42	1.2959%	17	0.6643%	210	8118%	20	0.8264%
35	1.0799%	17	0.6643%	200	7731%	17	0.7025%
32	0.9873%	16	0.6252%	190	7344%	17	0.7025%
32	0.9873%	15	0.5862%	180	6958%	16	0.6612%
31	0.9565%	15	0.5862%	180	6958%	16	0.6612%
31	0.9565%	15	0.5862%	180	6958%	16	0.6612%
30	0.9256%	13	0.5080%	160	6185%	16	0.6612%
29	0.8948%	13	0.5080%	150	5798%	15	0.6198%
28	0.8639%	12	0.4689%	140	5412%	14	0.5785%
28	0.8639%	11	0.4299%	120	4639%	12	0.4959%
28	0.8639%	11	0.4299%	120	4639%	12	0.4959%
27	0.8331%	11	0.4299%	120	4639%	12	0.4959%
27	0.8331%	11	0.4299%	110	4252%	11	0.4545%
26	0.7714%	11	0.4299%	110	4252%	10	0.4129%

Figure 6: Frequency order – FAC1501 English – IsiXhosa – IsiNdebele – IsiZulu

2.7. Alignment

The successful analysis of parallel texts depends on alignment. Alignment creates links between the ST and the TT. In the alignment process, the texts are matched at sentence level so that a sentence in the ST finds a corresponding sentence in the TT (cf. Figures 7 and 8).

English (Source Text)	IsiZulu (Target Text)
IN CONCLUSION: HELP WITH STUDY PROBLEMS	UKUPHETHA: USIZO MAYELANA NEZINKINGA EZIFUNDWENI ZAKHO
Your lecturers	Abafundisi bakho
Directorate: Counselling and Career Development	I-Directorate: Conselling and Career Development
Tutor programme	Uhlelo lwabasizikufundisa (ama-tutor)
Dear Student	Mfundi othandekayo
This tutorial letter contains technical advice on how to prepare assignments.	Le ncwadi yokufunda iqukethe iziboniso zokuthi uwenza kanjani ama-asayimenti.
In particular, it focuses on the skills required to produce an essay assignment of a high quality.	Ikakhulukazi, igxile emakhonweni afunekayo ukuze umfundi abhale i-asayimenti eyi-eseyi esezingeni eliphezulu.
It furthermore outlines a logical and organised approach to completing an essay assignment.	Iphinde futhi ichaze kafushane mayelana nendlela ephusile nehlelekile yokwenza i-asayimenti eyi-eseyi.
The quality of an assignment also depends on factors that are not discussed in this tutorial letter.	Ukuma kwe-asayimenti kuphinde kuncike kwezinye izinto ezingabalulwe kule ncwadi yokufunda.
Such factors include your depth of understanding of the topic in question, your ability to argue intelligently and to engage effectively with the literature, your ability to use ideas correctly, your originality and your awareness of the real world.	Lezi zinto zifaka phakathi ulwazi olujulile onalo ngesihloko okumele ubhale ngaso, ukwazi ukuchaza ngendlela enobuhlakani kanye nokusebenzisa imibhalo ebhalwe phambilini ngendlela ehlelekile, ukwazi ukusebenzisa imibono ngendlela efanele, ukusebenzisa imibono okungeyakho kanye nokubonisa ulwazi ngezinto ezenzeka emhlabeni.
However, to combine all these factors effectively, you should use the basics outlined in this tutorial letter.	Kodwa-ke, ukuze ukwazi ukuhlanganisa zonke lezi zinto ngendlela ehlelekile, kuzomele usebenzise imiyalelo okuchazwe ngayo kule

Figure 7: DVAALLD English-IsiZulu-aligned parallel texts

English (Source Text)	IsiZulu (Target Text)
ACADEMIC DISHONESTY	UKUNGATHEMBEKI EMFUNDWENI
Plagiarism	Ukukopela
Plagiarism is the act of taking the words, ideas and thoughts of others and presenting them as your own.	Ukukopela isenzo sokuthatha amagama, imibono nemicabango yabanye uyethule njengeyakho.
It is a form of theft which involves several dishonest academic activities, such as the following:	Kuyindlela yokweba efaka imisebenzi eminingana yokungathembeki yokufunda, njengokulandelayo:
Cutting and pasting from any source without acknowledging the source.	Ukusika nokunamathisela kunoma yimuphi umthombo ngaphandle kokuvuma ukusebenzisa umthombo wolwazi.
Not including or using incorrect references.	Ukungafaki noma ukusebenzisa imithombo yolwazi engalungile.
Paraphrasing without acknowledging the original source of the information.	Ukubeka amagama ngaphandle kokuvuma umthombo wangempela wolwazi.
Cheating	Ukukopela
Cheating includes, but is not limited to, the following:	Ukukopela kufaka phakathi, kepha akugcini kulokhu, okulandelayo:
Completing assessments on behalf of another student, copying from another student during an assessment or allowing a student to copy from you.	Ukubhala ukuhlolwa egameni lomunye umfundi, ukukopisha komunye umfundi ngesikhathi sokuhlolwa noma ukuvumela umfundi ukuthi akopishe kuwe.
Using social media (e.g. WhatsApp, Telegram) or other platforms to disseminate assessment information.	Kusetshenziswa izinsiza zokuxhumana (isib. WhatsApp, iTelegramu) noma amanye amapulatifomu okusabalalisa imininingwane yokuhlola.
Submitting corrupt or irrelevant files.	Ukuhambisa amafayela angasebenzi noma angahlobene nomsebenzi.
Buying completed answers from tutors or internet sites (contract cheating).	Ukuthenga izimpendulo ezibhaliwe ikumathutha noma kumasayithi e-inthanethi (ukukopela ngenkontileka).
More information about plagiarism can be downloaded on the link	Enye imininingwane mayelana nokukopela ingalandwa kwilinki

Figure 8: FAC1501 English-IsiZulu-aligned parallel texts

The alignment process is very important for the successful operation of the software. If the sentences are not aligned properly (as often happens), a menu of options makes it possible to split and merge sentences or segments (see Figure 9: note drop-down menu).

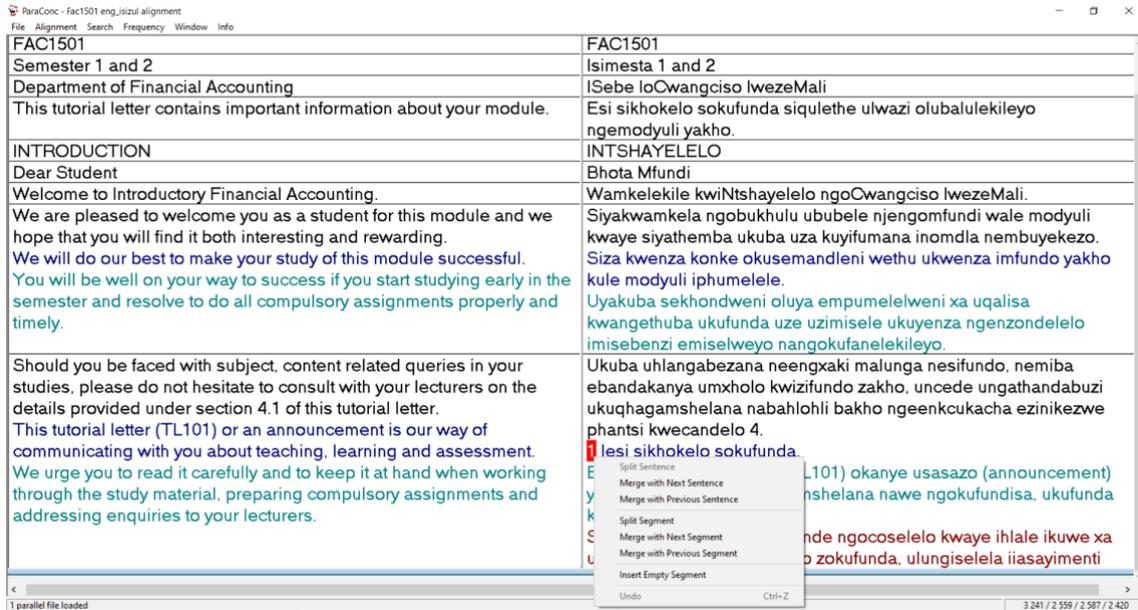


Figure 9: Alignment – split and merge sentences

3. DISCUSSION

One of the advantages of using a corpus is that researchers may access words in their real context (Moropa 2004). Context provides knowledge patterns that offer more information about the use of a term. This can only be possible by compiling a corpus of a specific domain in all languages. Such an exercise helps in defining equivalences. Moreover, researchers become empowered to coin words in their languages or use other word formation strategies. Table 2 shows the texts that were queried from human and accounting sciences.

STs	TTs and language	Size/words
Tutorial letter 301 for Honours in Development Studies (DVAALLD)		10,228
	<i>Izifundo Zentuthuko (IsiZulu)</i>	8,715
Tutorial letter 101 Introductory Financial Accounting (FAC1501)		3,241
	<i>Isikhokelo sokufunda 101 – Intshayelelo ngoCwangciso lwezeMali (IsiXhosa)</i>	2,559
	<i>Incwadi yokufundisa 101 – Introductory Financial Accounting (IsiNdebele)</i>	2,587
	<i>Incwadi yokufundisa 101 – Isethulo Sokubalisisa Izimali (IsiZulu)</i>	2,420
	<i>Incwadzi Yekufundzisa 101 – Singeniso Kutekuphatfwa Kwetimali (SiSwati)</i>	2,767

Table 2: The texts analysed in the study

When the texts are properly aligned, the researcher can search for a specific term and its translation equivalent using *ParaConc*. In this study, a keyword search for tutorial letter in the DVAALLD file was performed. The window showed two texts, the English text above and the IsiZulu translation below. The search for the English word ‘tutorial’ (Figure 10) had nine hits: the keyword is highlighted in blue and its collocates in red.

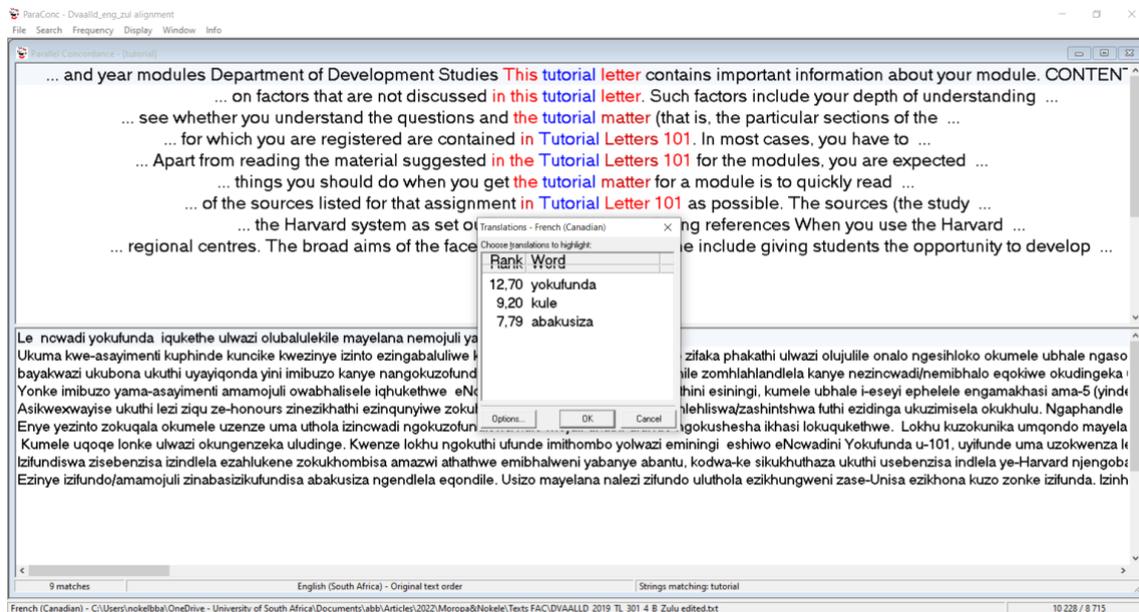


Figure 10: Identifying translation equivalents for tutorial letter in IsiZulu (DVAALLD)

To find the translation equivalent of ‘tutorial’, we right-clicked on the IsiZulu window, and a drop-down menu with options appeared. We then scrutinised the results and found that the equivalent for IsiZulu in the DVAALLD file is *incwadi yokufunda*. Similar results were found in the FAC1501 file: IsiNdebele *incwadi yokufundisa*, IsiZulu *incwadi yokufunda*, and SiSwati *tincwadi lekufundzisa*. IsiXhosa showed a different result which was *isikhokelo*. Although the translators had different solutions for ‘tutorial letter’: *incwadi yokufundisa/lekufundzisa* (‘a book for teaching’), *incwadi yokufunda* (‘a book for learning’) and *isikhokelo* (‘a guide’), they all essentially imply the same knowledge information that involves learning.

A search for ‘trial balance’ (FAC1501) yielded four results and the translation equivalents in SiSwati were paraphrases such as *simo semabhuku etimali* and *simo setimali*. IsiNdebele used indigenisation *ithrayalibhalans* (‘trial balance’), while IsiXhosa settled for *ibhalansi yolingo*, which is the equivalent for ‘general ledger’ (cf. Figure 11), another common term in accounting. IsiNdebele and IsiXhosa translators opted for *ileja*, an indigenisation strategy throughout the texts, while IsiZulu and SiSwati translators used

paraphrasing: in the case of IsiZulu *ibhuku elijwayekile lemali* ('a book that is common for finance / cash'), *ibhukwana* ('a small book'), *ibhukwana elijwayekile* ('a small book that is common'), *ibhukwana elivamile* ('a book that is known') and, in the case of SiSwati, *umculu wekubikwa kwemali* ('a book that records money/finances'), *umculu wemarikhodi etetimali* ('book for financial records'), and *kumarekhodi embiko wetetimali* ('in the record of financial report').

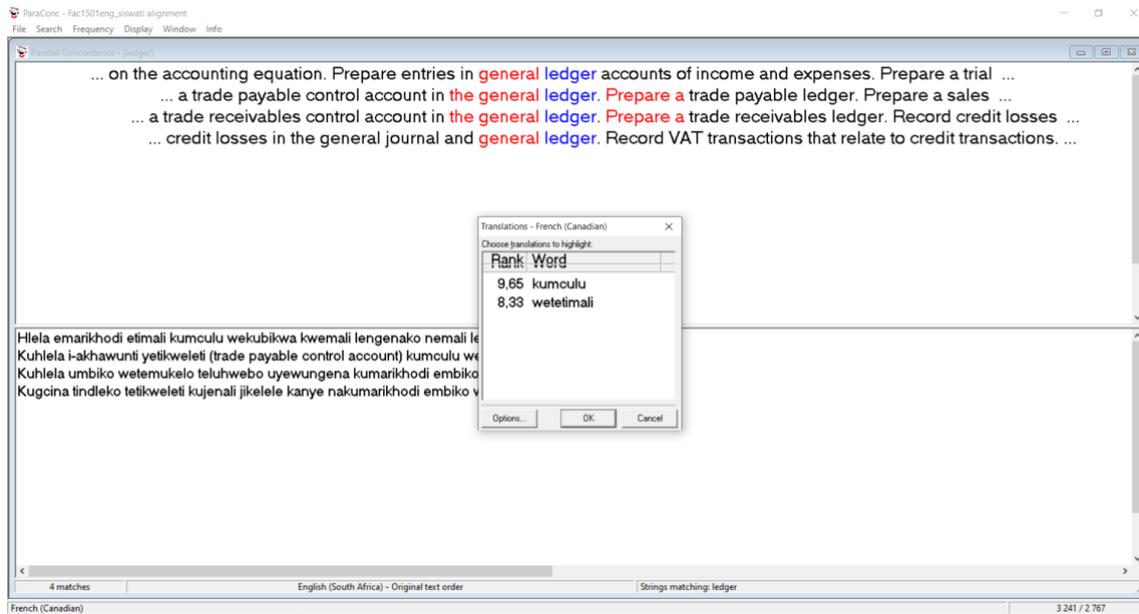


Figure 11: Identifying translation equivalent for general ledger in SiSwati (FAC1501)

Another finding was the translation of the word 'assignment' in IsiNdebele as *umtlolophenyo* and *sivivinyo* in SiSwati, which are coined or derived words, while the IsiXhosa and IsiZulu used indigenisation *iasyimenti/iasayinimenti*. Earlier on, we pointed out that a corpus can be used as resource from which terms can be drawn. For example, if IsiXhosa does not have a term for 'assignment', as illustrated earlier, and does not make use of a borrowed word, the researcher can consult the other sister languages in the Nguni language family and borrow it from them. In this instance, IsiXhosa translators can learn from IsiNdebele, which opted for the indigenous term *umtlolophenyo*, or they can coin their own IsiXhosa word related or similar to the IsiNdebele word. *Umtlolophenyo* is a compound word which is derived from *umtlolo* ('writing') and *phenyo* ('research'). If the IsiXhosa translators decide to learn from IsiNdebele, they may use compounding and form *ubhalophando* as an equivalent for 'assignment'. This way a new term may be formed as a synonym. For 'journal' and 'receipts', borrowing was used and indigenisation of the loan word was adopted, that is, the term was spelt according to the orthographical

rules of the borrowing language: IsiXhosa *ijenali*, (*yee*)*risithi*, SiSwati *ijenali*, and IsiNdebele *ijenali* and (*yama*)*risidi*.

Another useful tool in *ParaConc* is the hot word tool, which helps in identifying both possible translations and the knowledge patterns such as collocations or synonyms. Hot words are selected by looking at the frequency of words. The top-ranking words may include translations, translations of collocations, and collocations of the search word. This is illustrated in Figure 12, which shows the occurrences of ‘accounting’ in the English-SiSwati alignment.

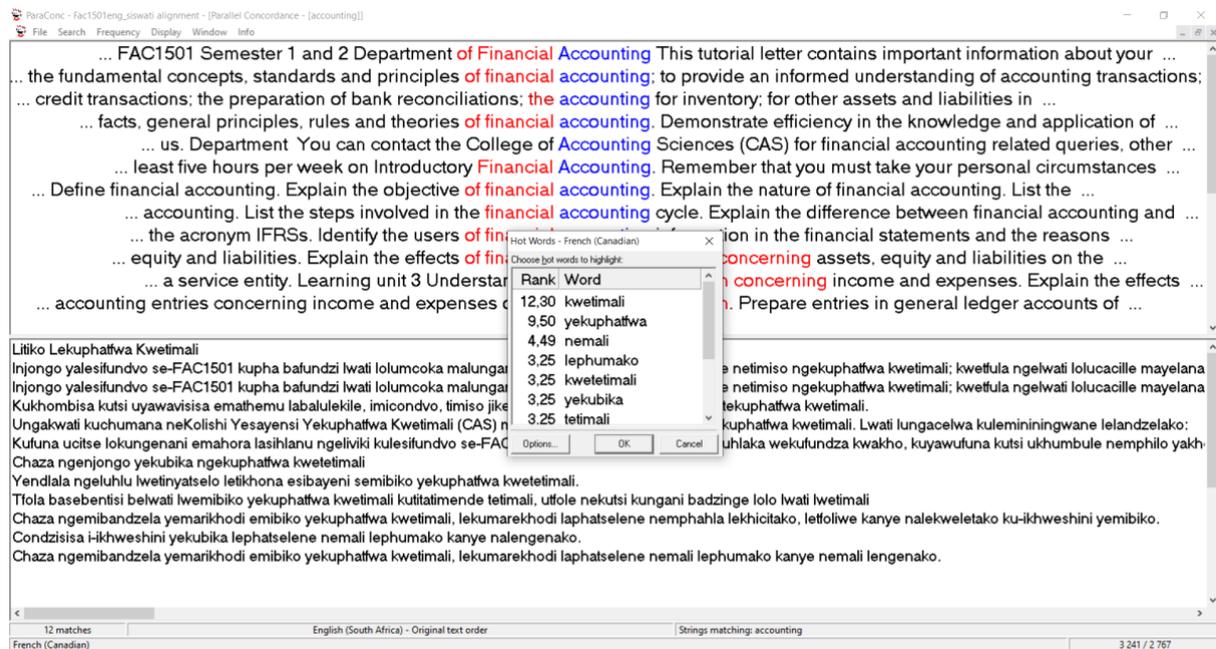


Figure 12: Hot words for ‘accounting’ (FAC1501-English - SiSwati)

The search revealed several hot words suggesting SiSwati equivalents, of which *kwetimali* (12.30) and *yekuphatfwa* (9.50) ranked the highest. Results for IsiNdebele revealed ‘accounting’ as a loan word, which suggests that most of the time accounting was reproduced and not translated. ‘Financial’ with a co-occurring word, as in ‘financial accounting’, was also not translated. The high frequency words for IsiXhosa were *lwezemali*, *lwemali*, and *mali*, which suggests that accounting in IsiXhosa has to do with money. The information revealed for IsiZulu was *kwezimali* and *ezibalweni*. When comparing the four languages it seems clear that IsiXhosa, IsiZulu and SiSwati conceptualise ‘accounting’ in the same way, because the highest-ranking words are *lwezemali*, *kwezimali* and *kwetimali*, respectively. In this sense, IsiNdebele can use the same translation strategy instead of using the loan word(s).

Other terms that were queried were generic terms attested in all level-one tutorial letters. These terms were ‘module’, ‘electronic reserves’, ‘internet’, ‘online’, and ‘plagiarism’. Similarity in the translation of certain words was observed during our analysis. ‘Module’ was translated as *imojuli* in IsiZulu and IsiNdebele, and *imodyuli* in IsiXhosa. The only difference between these words is in spelling and pronunciation. The translation strategy is the same in all three languages, that is, indigenisation. The equivalent for SiSwati is *tifundvo* (‘lesson’), which is different from other sister languages. SiSwati opted for an indigenous term that carries a similar meaning. *Imojuli/imodyuli* and *tifundvo* can then be taken as referring to the same term.

‘Electronic reserves’, ‘internet’, and ‘online’ are among the terms that generally pose a challenge to translators of African languages. The analysis of the five texts revealed that translators sometimes use ‘internet’ and ‘online’ interchangeably. IsiXhosa and IsiZulu, for instance, opt for *ku-inthanethi* or *ngeintanethi*, while IsiNdebele and SiSwati use *nge-online/ya-online/ta-online*. IsiXhosa sometimes uses *ngomoya*, which refers to waves rather than wind. This shows that translators understand how the ‘internet’ or ‘online’ terms work. The hot word tool revealed that the use of ‘online’ in the SiSwati texts has a ranking of 7.21, while *nge-online* in IsiNdebele has a ranking of 6.26, and *ku-inthanethi* has a ranking of 6.71 in the context where ‘online’ is used. These rankings suggest that the words ‘internet’ and ‘online’ are treated as synonyms in these languages and can therefore be incorporated in their lexicon.

‘Plagiarism’ is another term that is difficult to translate into African languages. Most of the time it is paraphrased in long sentences, as in *ukuthatha umsebenzi womnye umntu uwenze owakho* (‘to take one’s work and present it as your own’), in IsiXhosa, or *ubunikazi bomsebenzi ekungasiwo wakho i-plagiarism* (‘ownership of work that is not your own’) in IsiNdebele. IsiZulu and SiSwati simplify this explanation by opting for *ukukopela* (‘copying’) or *kukopa* (‘cheating’). It is interesting to note that there is a kind of uniformity among the languages regarding the translation of academic dishonesty: *ukungathembeki emfundweni* (‘to be dishonest in education’) in IsiZulu and IsiNdebele, *kungetsembeki emsebenzini* (‘to be dishonest in your work’) in SiSwati, *ungathembeki kwimfundo* (‘to be dishonest in education’) in IsiXhosa. There is also curriculum transformation: *ukuguqulwa kwekharikhulamu* (‘changing in the curriculum’) in IsiZulu, *ukutjhugululwa kwekharikhyulamu* (‘changing in the curriculum’) in IsiNdebele, *tingucuko kukharikhulamu* (‘changing in the curriculum’) in SiSwati, and *Utshintsho*

kwikharityhulamu ('changing in the curriculum') in IsiXhosa. Such similarities confirm the acceptance of the terms or concepts by the community.

The analysis has also revealed that a corpus can be beneficial in terminology development when looking at the knowledge information and knowledge patterns displayed in the texts. Researchers and translators can retrieve information from such knowledge patterns when developing terms in their own languages. Another advantage of using a corpus is that researchers can see the translation strategies adopted by other language practitioners when dealing with equivalence or lack of terminology in their own languages. This then puts them in a better position to make informed decisions.

The process explained above illustrates how a parallel corpus can be used to extract terms and their translations, thereby enriching the indigenous languages. Corpora as authentic resources provide terms in use. The frequency of use also indicates whether the term has been accepted by the speech community. The similarity that is observable in how IsiNdebele, IsiXhosa, IsiZulu, and SiSwati translators deal with specialised terminology indicates that these languages are being developed and also shows that related languages can borrow from each other to solve the problem of lack of terms.

The step towards corpus building and use of software to query data will improve the process of the identification of terms, both in monolingual and parallel platforms. In dictionaries, information about frequency or generalisation of use is not provided in a consistent manner, whereas such facts can be obtained somewhat more readily from parallel texts. As Bowker (2000: 21) states, parallel corpora contain a range of terms that is wider than dictionaries, present terms in context, and are more current than dictionaries. They allow translators to acquire both specialised conceptual and linguistic knowledge about terms and, as they are available in electronic form, they can increase the scale and speed of a translator's research. Teubert (2005: 98) concludes that:

many words in our ordinary language have unspecific meanings which cannot be described without referring to the context in which they occur, but this is what dictionaries, due to their constraints in space, cannot do effectively.

The meaning of words is created in texts and this renders specialised status to terms, which is why it is fundamental to identify the context where texts are produced and to describe the texts that constitute the corpus (Moreira 2014).

4. CONCLUSION

Parallel texts present terms in authentic contexts allowing terminologists and translators to acquire specialised conceptual and linguistic knowledge. The extraction of terminology from parallel corpora (considering the different language pairs) is not only feasible but also extremely useful in developing scientific and technical vocabulary for large-scale use in Unisa. In addition, it is worth noting that the researcher, that is, the linguist, verifies the list of terms. History shows that the indigenous African languages are capable of drawing on their own resources and can create the necessary terms from their own vocabulary.

To reinforce language transformation within Unisa, corpus compilation and the use of software tools should be part of the agenda of language development for teaching, learning, and research. This is in line with the language policy framework that encourages institutions of Higher Learning in South Africa to develop language plans and strategies that will enhance the development and promotion of indigenous African languages as centres of research and scholarship. By constructing corpora as resources for linguistic research, Unisa will also be acknowledging the authors of the texts, as well as the language practitioners who translate, edit, and proofread the texts that are public. Unisa is the only university in South Africa that, in its transformation agenda, has committed to build capacity for all official South African languages. Furthermore, it is the only university in South Africa that teaches all official South African languages. Electronic written corpora should serve as a tool for the development of South African languages, given that a corpus provides the researcher with a wealth of linguistic data instantly. Corpus-based linguistic research within the University of South Africa will transcend terminology development because the corpus-based approach can be applied to empirical investigations in almost any discipline with its attendant linguistic challenges.

REFERENCES

- Alexander, Neville. 2003. *African Renaissance and the Use of African Languages in Tertiary Education*. Cape Town: The Estate of Neville Edward Alexander.
- Baker, Mona Baker. 1995. Corpora in translation studies: An overview and some suggestions for future research. *Target* 7/2: 223–243.
- Barlow, Michael. 2008. *ParaConc and Parallel Corpora in Contrastive and Translation Studies*. Houston: Athelstan.

- Bowker, Lynne. 2000. Towards a methodology for exploiting specialized target language corpora as translation resources. *International Journal of Corpus Linguistics* 5/1: 17–52.
- Department of Education. 2002. *Language Policy for Higher Education*. Pretoria: Government Printers.
- Department of Education. 2020. *Language Policy for Higher Education*. Pretoria: Government Printers.
- De Schryver, Gilles-Maurice and Jacobus Daniëlle. 2005. Managing eleven parallel corpora and the extraction of data in all official South African languages. In Walter Daelemans ed. *Multilingualism and Electronic Language Management*. Pretoria: Van Schaik, 100–122.
- Gauton, Rachéle and Gilles-Maurice De Schryver. 2004. Translating technical texts into Zulu with the aid of multilingual and/or parallel corpora. *Studies in the Languages of Africa* 35/1: 148–161.
- Granger, Sylviane. 1998. *Learner English on Computer*. London: Longman.
- Khumalo, Langa, Valentine Azom and Peter Olukanmi. 2019. The design and implementation of a corpus management system for IsiZulu National Corpus. In Martin Doerr, Oyvind Eide, Oddrun Gronvik and Bjorghild Kjelsvik eds. *Humanists and the Digital Toolbox*. Oslo: Novus Forlag, 179–196.
- Kilgarriff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý and Vít Suchomel. 2014. The Sketch Engine: Ten years on. *Lexicography* 1/1: 7–36.
- Leech, Geoffrey. 1998. Learner corpora: What they are and what can be done with them. In Sylviane Granger ed. *Learner English on Computer*. London: Longman: xiv–xx.
- Madiba, Mbulungeni. 2001. Towards a model for terminology modernisation in the African languages of South Africa. *Language Matters* 32/1: 53–77.
- Madiba, Mbulungeni. 2004. Parallel corpora as tools for developing indigenous languages of South Africa. *Language Matters* 35/1: 133–147.
- Mlambo, Respect, Nomsa Skosana and Muzi Matfunjwa. 2021. The extraction of terminology list using ParaConc for creating a quadrilingual dictionary. *Southern African Linguistics and Applied Language Studies* 39/1: 82–91.
- Moreira, Adonay. 2014. A methodology for building a translator and translation oriented terminological resource. *InTralinea Online Translation Journal*. <https://www.intralinea.org/specials/article/2032>
- Moropa, Koliswa. 2004. A parallel corpus as a terminology resource for Xhosa: A study of strategies used to translate financial statements. *Language Matters* 35/1: 162–178.
- Moropa, Koliswa. 2005. *An Investigation of Translation Universals in a Parallel Corpus of English-Xhosa Texts*. Pretoria: University of South Africa dissertation.
- Moropa, Koliswa. 2007. Analysing the English-Xhosa parallel corpus of technical texts with ParaConc: A case study of term formation processes. *South African Linguistics and Applied Language Studies* 25/2:183–205.
- Moropa, Koliswa and Feziwe Martha Shoba. 2017. Language and terminology development in IsiXhosa: A history. In Russell H. Kaschula, Pamela Maseko and H. Ekkehard Wolff eds. *Multilingualism and Intercultural Communication: A South African Perspective*. Johannesburg: Wits University Press, 76–91.
- Ndhlovu, Ketiwe. 2012. *An Investigation of Strategies Used by Ndebele Translators in Zimbabwe in Translating HIV/AIDS Texts: A Corpus-based Approach*. Alice: University of Fort Hare dissertation.

- Ndhlovu, Ketiwe. 2016. Using ParaConc to extract bilingual terminology from parallel corpora: A case of English and Ndebele. *Journal of Literary Criticism, Comparative Linguistics and Literary Studies* 37/2: 1–12.
- Prinsloo, Daniël Jacobus. 1991. Towards computer assisted word frequency studies in Northern Sotho. *South African Journal of African Languages* 11/2: 54–60.
- Republic of South Africa. 1996. *The Constitution of the Republic of South Africa* (Act 108 of 1996). Pretoria: The Government Printer.
<https://www.gov.za/documents/constitution-republic-south-africa-1996>
- Sinclair, John. 1995. Corpus typology: A framework for classification. In Gunnel Mechers and Beatrice Warren eds. *Studies in Anglistics. Acta Universitatis Stockholmienses*. Stockholm: Almqvist & Wicksell, 17–33.
- Shoba, Feziwe Martha. 2018. *Exploring the Use of Parallel Corpora in the Compilation of Specialized Bilingual Dictionaries of Technical Terms: A Case Study of English and IsiXhosa*. Pretoria: University of South Africa dissertation.
- Teubert, Wolfgang. 2005. Language as an economic factor: The importance of terminology. In Geoffrey Barnbrook, Pernilla Danielsson and Michaela Mahlberg eds. *Meaningful Texts: The Extraction of Semantic Information from Monolingual and Multilingual Corpora*. London: Continuum, 96–106.
- University of South Africa. 2016. *Unisa Language Policy*.
<https://www.unisa.ac.za/policies>

Corresponding author

Koliswa Moropa

University of South Africa

E-mail: koliswa04@gmail.com

Department of Tuition Support and Facilitation of Learning

Muckleneuk Campus

1 Preller Street

Pretoria 0002

South Africa

received: March 2023

accepted: May 2023

APPENDIX

Appendix 1: Terms extracted from the corpus

ST term	TT term	Language	Translation strategy
Tutorial letter	<i>Incwadi yokufunda</i>	IsiZulu	Paraphrase
	<i>Incwadi yokufundisa</i>	IsiNdebele	Paraphrase
	<i>Tincwadi lekufundzisa</i>	SiSwati	Paraphrase
	<i>Isikhokelo</i>	IsiXhosa	Substitution
Trial balance	<i>Ibhalansi yomzamo</i>	IsiZulu	Paraphrase
	<i>Simo semabhuku etimali</i>	SiSwati	Paraphrase
	<i>simo setimali</i>		
	<i>Ithrayalibhalansi</i>	IsiNdebele	Indigenisation
	<i>Ibhalansi yolingo</i>	IsiXhosa	Paraphrase
General ledger	<i>Ibhuku elijwayekile</i>	IsiZulu	Paraphrase
	<i>lemali/</i>		
	<i>Ibhukwana</i>		
	<i>elijwayelekile/</i>		
	<i>ibhukwana elivamile</i>		
	<i>Ileja</i>	IsiNdebele	Indigenisation
	<i>Umculu wekubikwa</i>	SiSwati	Paraphrase
	<i>kwemali/</i>		
	<i>umculu wemarikhodi</i>		
	<i>etetimali /</i>		
	<i>kumarekhodi embiko</i>		
	<i>wetetimali</i>		
	<i>Ileja</i>	IsiXhosa	Indigenisation
Assignment	<i>Iasayinimenti</i>	IsiZulu	Indigenisation
	<i>Umtlolophenyo</i>	IsiNdebele	Substitution
	<i>Sivivinyo</i>	SiSwati	Substitution
	<i>Iasayimenti</i>	IsiXhosa	Indigenisation
Journal (of receipts)	<i>Ijenali (yamarisidi)</i>	IsiZulu	Indigenisation
	<i>Ijenali yamarisidi</i>	IsiNdebele	Indigenisation
	<i>Ijenali</i>	SiSwati	Indigenisation
	<i>ijenali (yeerisithi)</i>	IsiXhosa	Indigenisation
Module	<i>Imojuli</i>	IsiZulu	Indigenisation
	<i>Imojuli</i>	IsiNdebele	Indigenisation
	<i>Tifundvo</i>	SiSwati	Substitution
	<i>Imodyuli</i>	IsiXhosa	Indigenisation
Accounting	<i>Ukubalwa kwezimali</i>	IsiZulu	Paraphrase
	<i>Ukubalwa kwemali</i>	IsiNdebele	Paraphrase
	<i>Ubalo lwezimali</i>	SiSwati	Paraphrase
	<i>ucwancisomali</i>	IsiXhosa	Substitution
Financial accounting	<i>Ukubalwa kwemali</i>	IsiZulu	Paraphrase
	<i>yokubalwa kwemali</i>		
	<i>wokubalwa kwemali</i>		
	<i>Financial accounting</i>	IsiNdebele	Loan word
	<i>Ukuphatfwa kwetimali</i>	SiSwati	Paraphrase
	<i>Ucwancisomali</i>	IsiXhosa	Substitution
Inventory control	<i>Ulawulo lokusungula</i>	IsiZulu	Substitution
	<i>Ulawulo lwenani yepahla</i>	IsiNdebele	Paraphrase
	<i>Luhlelo loluchaphaluhlu</i>	SiSwati	Paraphrase
	<i>Lwempahla</i>		
	<i>Ulawulo lwempahla</i>	IsiXhosa	Substitution

ST text term	TT text term	Language	Translation strategy
Online	<i>Ku-inthanethi</i>	IsiZulu	Indigenisation
	<i>Nge-online</i>	IsiNdebele	Loan word
	<i>Online</i>	SiSwati	Loan word
	<i>Ya-online</i>		
	<i>Ta-online</i>		
	<i>Ngomoya</i> <i>Ngeintanethi</i>	IsiXhosa	Substitution, indigenisation
Purpose	<i>Inhloso</i>	IsiZulu	Substitution
	<i>Umnqopho</i>	IsiNdebele	Substitution
	<i>Injongo</i>	SiSwati	Substitution
	<i>Injongo</i>	IsiXhosa	Substitution
Outcomes	<i>Imiphumela</i>	IsiZulu	Substitution
	<i>Imiphumela</i>	IsiNdebele	Substitution
	<i>Imiphumela</i>	SiSwati	Substitution
	<i>Iziphumo</i>	IsiXhosa	Substitution
Academic dishonesty	<i>Ukungathembeki</i> <i>emfundweni</i>	IsiZulu	Paraphrase
	<i>Ukungathembeki</i> <i>emfundweni</i>	IsiNdebele	Paraphrase
	<i>Kungetsembeki</i> <i>emsebentini</i>	SiSwati	Paraphrase
	<i>Ukungathembeki</i> <i>kwimfundo</i>	IsiXhosa	Paraphrase
Electronic reserves	<i>Imithombo eku-inthanethi</i>	IsiZulu	Paraphrase
	<i>Imithombo eyi-Electronic reserves (e-reserves)</i>	IsiNdebele	Paraphrase plus loan word
	<i>lwati lolugcinwe emishinini</i>	SiSwati	Paraphrase
	<i>izixhobo ezigcinwe ngoomatshini</i>	IsiXhosa	Paraphrase
Assessment criteria	<i>Indlela yokuhlola</i>		Substitution
	<i>Amaqhingha wokuhlola</i>		Paraphrase
	<i>Indlela yekuhlolwa kwebafundzi</i>		Paraphrase
	<i>Iindlela zovavanyo</i>		Substitution
Curriculum transformation	<i>Ukuguqulwa</i>	IsiZulu	Substitution
	<i>kwekharikhulamu</i>		
	<i>Ukutjhugululwa</i>	IsiNdebele	Substitution
	<i>kwekharikhyulamu</i>		
	<i>tingucuko kukharikhulamu</i>	SiSwati	Substitution
	<i>Utshintsho</i> <i>kwikharithulamu</i>	IsiXhosa	Substitution
Resources	<i>izinsizakusebenza</i>	IsiZulu	Substitution
	<i>Imithombo</i>	IsiNdebele	
	<i>Tinsita tekufundza resources</i>	SiSwati	Paraphrase
	<i>Izibonelelo zokusebenza</i>	IsiXhosa	Paraphrase
Log on	<i>Ngena</i>	IsiZulu	Substitution
	<i>Loga</i>	IsiNdebele	Indigenisation
	<i>Condza</i>	SiSwati	Substitution
	<i>Ngena</i>	IsiXhosa	Substitution

ST text term	TT term	Language	Translation strategy
Define	<i>Chaza</i>	IsiZulu	Substitution
	<i>Hlathulula</i>	IsiNdebele	Substitution
	<i>Chaza</i>	SiSwati	Substitution
	<i>Chaza</i>	IsiXhosa	Substitution
Identify	<i>Thola</i>	IsiZulu	Substitution
	<i>Khomba</i>		
	<i>Tjengisa</i>	IsiNdebele	Substitution
	<i>Tjola</i>	SiSwati	Substitution
	<i>Chonga</i>	IsiXhosa	Substitution
	<i>Bonisa</i>		
Explain	<i>Chaza</i>	IsiZulu	Substitution
	<i>Hlathulula</i>	IsiNdebele	Substitution
	<i>Chaza</i>	SiSwati	Substitution
	<i>Cacisa</i>	IsiXhosa	Substitution
Plagiarism	<i>Ukukupela</i>	IsiZulu	Indigenisation
	<i>Ubunikazi bomsebenzi ekungasiwo wakho</i>	IsiNdebele	Paraphrase plus loan word
	<i>i-plagiarism</i>		
	<i>Kukopa</i>	SiSwati	Indigenisation
	<i>Ukuthatha umsebenzi womnye umntu uwenze owakho</i>	IsiXhosa	paraphrase
Development Studies	<i>Izifundo zezentuthuko</i>	IsiZulu	Substitution
Critical reading	<i>Ukufunda</i>	IsiZulu	Paraphrase
	<i>ngokucubungula</i>		
References	<i>Ukukhonjiswa kolwazi oluthathwe emibhalweni yabanye abantu</i>	IsiZulu	Paraphrase
Bibliographic details	<i>Imininingwane yokushicilelwa</i>	IsiZulu	Substitution
Online assignments	<i>Ama-asayimenti atyelwa ngamakhompyutha (athunyelwa online)</i>	IsiZulu	Paraphrase
Critically discuss	<i>Xoxa ngokucubungula</i>	IsiZulu	Substitution
Distinguish	<i>Yehlukanisa</i>	IsiZulu	Substitution
Framework	<i>Uhlaka</i>	IsiZulu	Substitution