TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF GRAPHS

ABBREVIATIONS AND ACRONYMS

The following abbreviations and acronyms are used in this study:

AFS:	Annual Financial Statements
CTS:	Corpus-based Translation Studies
DAC:	Department of Arts and Culture
DTS:	Descriptive Translation Studies
EngXhPC:	English-isiXhosa Parallel Corpus
FSB:	Financial Services Board
GDX:	The Greater Dictionary of Xhosa/isiXhosa
ISX:	IsiChazi-magama sesiXhosa
LSP:	Language for Special Purposes
LGP:	Language for General Purposes
MWU/E:	Multiword Units/Expressions
PanSALB:	Pan South African Language Board
SC:	Subject Concord
SL:	Source Language
ST:	Source Text
TT:	Target Text
TL:	Target Language
XDP:	Xhosa Dictionary Project
XNLU:	IsiXhosa National Lexicography Unit

CHAPTER 1

Introduction

1.1 Background to and rationale behind the research problem

In recognition of the historically diminished use and status of indigenous languages of South Africa, Section 6 (2) of the Constitution of the Republic of South Africa, Act 108 of 1996 mandated the state 'to take practical and positive measures to elevate the status and the use of indigenous languages'. This led to the implementation of functional multilingualism which in turn increased the demand for translating technical texts from either English or Afrikaans into various indigenous languages. However, to embark on the translation process, one needs terminological resources such as glossaries and bilingual dictionaries but these are unfortunately lacking in many African languages. In the overview of isiXhosa lexicography, bilingual lexicography dates back as 1872 (cf. Pahl et al 1989; Mtuze, 1992; Moropa and Kruger, 2000). Of the bi/multilingual dictionaries that were published in the two decades the majority are general-purpose dictionaries; amongst these are: the *Oxford English-Xhosa Dictionary*, 1985, by Fischer, Weiss, Mdala and Tshabe, *The Greater Dictionary of Xhosa* (Volume 3 Q-Z) by Pahl, Pienaar, and Ndungane, 1989, *The Greater Dictionary of isiXhosa*, Volume 2 (K-P) 2003, by Mini, Tshabe, Shoba and van der Westhuizen, with Volume 1 (A-J) first published in 2006 by Tshabe and Shoba, as well as *Isichazi-magama seMathematika neNzululwazi 2014, Oxford IsiXhosa-English English-isiXhosa School Dictionary* by De Schryver and Reynolds, 2014. This problem is more apparent in specialised bilingual dictionaries in the fields of commerce, science and technology to be used as resources by translators of special field texts. Moropa (2004) and Chabata (2013) explain that the lack of specialised terminology is attributed to the underdevelopment and underutilisation of indigenous languages as languages of intellectualisation. In the examples cited above there is only one bilingual specialised dictionary, namely, *Isichazi-magama seMathematika neNzululwazi* (2014) whose aim is to aid school children to understand the difficult concepts of Maths and Science. *The Oxford Bilingual School Dictionary* published in 2014 was the first corpus-based dictionary in the history of isiXhosa bilingual lexicography. It only targets school

children. This indicates that specialised bilingual dictionaries that focus on specific domains are still lagging behind in isiXhosa.

The general failure to use African languages in various specialised fields has hindered their development and growth in a language for specific purposes (LSP) vocabulary. This view is shared by Klein (2010) who observes:

> The nine official African languages still lack the status of English and Afrikaans. One reason for this is that they are still underdeveloped in the area of electronic, online and specialised dictionaries and Human Language Technology.

In response to this challenge, this study explores the use of parallel corpora as a means of compiling specialised dictionaries of technical terms in African languages with particular reference to English and isiXhosa in order to address the lack of resources in the Xhosa language. A technology based methodology was also selected because it is faster and more accurate compared to the traditional method.

In South Africa, a number of researchers have created parallel corpora for various purposes; nevertheless, the existence of these corpora intimates that skill and knowledge on corpus building are available, which makes this methodology feasible. Examples of corpus-based research that has been undertaken in translation studies in South Africa include: Gauton and De Schryver (2004); Moropa (2004, 2005, 2007); Madiba (2004); Ndlovu (2009). Madiba (2004) designed a Special Language Corpora for African Languages (SPeLCaL) and demonstrated the two critical effects of corpora as being:

- To provide a language resource for research into 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).

Moropa (2004) demonstrates how a computer-based bilingual corpus of documents translated from English into isiXhosa could serve as a terminology resource for Xhosa translators. An English-Xhosa sub-corpus of financial terms was used and language-based analysis software called ParaConc was utilised to query frequency and concordance lists. She also compared the information recorded in bilingual dictionaries

of English-isiXhosa with that retrieved from the parallel corpora. Moropa's study is similar to the current study in that both utilise financial texts in corpus building, but nonetheless the approaches used to analyse data are different. Whilst Moropa focused on strategies for term creation, this study will explore how ParaConc can be used to extract bilingual terminology from a parallel corpus for dictionary making. This is done in an attempt to design improved dictionaries that could satisfy the needs of different potential users such as translators, semi-specialists, specialist and general users. Additionally, the development of bilingual specialised dictionaries should contribute to the standardisation and growth of African languages as languages of technology, commerce and science. There is a dire need for specialised dictionaries in African languages. The following are the viewpoints in this regard.

The necessity of specialised dictionaries in African languages is affirmed by a notable number of African languages scholars (cf. Chabata 2013; Nkomo, 2010; Gouws and Prinsloo, 2005). In affirming the role of specialised dictionaries, Nkomo (2010:377) argues that 'LSP dictionaries in indigenous African languages need to be seen as potentially useful tools and resources which may solve problems faced in the development and acquisition of specialised languages as well as the translation of specialised texts'. Chabata (2013) regards the compilation of specialised dictionaries as one of the major steps in raising the status of the African languages in preparation for their promotion to high profile functions in society. The discussions by these two scholars support the aim of the current researcher. The claim that African languages cannot be used in special domains should be history.

This study which is interdisciplinary in approach (translation and lexicography) will hopefully contribute to (a) knowledge on corpus-based lexicography in African languages, (b) the standardisation of specialised terminologies, (c) resource development by showing how specialised bilingual dictionaries can be developed from parallel corpora and (d) the development of user-friendly dictionaries that will serve a variety of users such as translators, translation students, language professionals, second language speakers, subject specialists and the general public.

1.2 Statement of the problem

The traditional method of dictionary making, which involves collecting terms by means of interviews, literature, newspapers and translating them, is a long and laborious process, which at times takes years of hard work and a lot of financial backing. The compilation of *The Greater Dictionary of Xhosa* (Volume 3 Q-Z), for example, commenced in 1968 and was published in 1989, approximately more than twenty years later. The methods of capturing data relied on previous dictionaries, perusal of literature and manual editing. The method was manual and labour intensive. This method has led to few dictionaries being produced in African languages, calling for a more efficient and faster process. This study thus explores how parallel corpora can be used as a basis for the compilation of English–isiXhosa specialised dictionaries of financial terms. The study will address the following research questions:

1) How can ParaConc, a parallel concordancer, be used to extract bilingual information from parallel corpora?
2) How can information extracted from parallel corpora be used to create bilingual dictionaries?

In order to fulfil both aims, a parallel English-isiXhosa corpus will be created and this will be covered in Chapter 3 which deals with the methodology of this study. Parallel corpora comprise of data from two languages which makes it suitable for the creation of bilingual dictionaries. This study utilises translation methodologies embedded within Corpus-based Translation Studies to inform the field of lexicography and more details are provided in the following sections.

1.3 Aims and objectives of the study

In line with the above stated questions, the **main aim** of this study is to investigate how specialised parallel corpora can be used in the compilation of English-isiXhosa bilingual dictionaries of financial terms.

The objectives of the study therefore are:

- To create an English-isiXhosa Parallel Corpus of financial texts
- To extract bilingual terminology for dictionary making from the parallel corpus
- To illustrate how information extracted from parallel corpora can be used to address communicative and cognitive lexicographic functions that will meet the needs of different users of specialised dictionaries.

It is therefore hoped that the achievement of the aim and the objectives will not only improve the structure of specialised dictionaries but will also increase relevant specialised terminology in African languages, with particular reference to isiXhosa, and promote them as languages of science and technology.

1.4 Methodology

This section presents the methods used to collect and analyse data in this study. The researcher commences by presenting a theoretical framework followed by the research type or design, data collection tools and procedures, data analysis and interpretations used in this study. Lastly, the key terms and concepts that are fundamental in the current study are defined. In the study, three theories are applied: corpus-based translation studies, Wiegand's general theory of lexicography and the theory of lexicographic functions. These theories are applied in a complementary manner and are explained in brief below.

1.5 Theoretical framework

According to Kerlinger (1986:9) a theoretical framework refers to ‘a set of interrelated constructs or concepts, definitions and propositions that present a systematic view of phenomenon by specifying relations among variables with the purpose of explaining and predicting phenomena’. A theoretical framework forms the basis of the research problem. This study is methodological in approach in that it explores how ParaConc can be used to extract bilingual information from parallel corpora for the compilation of bilingual dictionaries. Corpus building in translation studies falls under the Corpus-based Translation Studies methodology and this framework will be used to guide the study. Corpus-based translation studies (CTS) is a relatively new area of research which emerged in the late 1990’s and is informed by corpus linguistics. CTS is now recognised as a major paradigm that has transformed analysis within the discipline of translation studies. It is relevant in the current study, because the use of corpora is at the centre. It shows a shift from prescriptive analysis of texts and emphasises the description of the smallest details in the corpus research. CTS develops from corpus linguistics.

Baker (2010:93) describes corpus linguistics as a popular field of linguistics which involves the analysis of very large collections of electronically stored texts, aided by computer software. In the current study, the researcher will collect and analyse a parallel corpus that constitutes a variety of translated texts of financial terms so that they can be easily manipulated by means of specific corpus analysis tools and software. Corpus linguistics displays salient features that are going to be beneficial for the proposed study. Teurbet (2001:129) observes that corpus linguistics has combined three different approaches, namely, the identification of language data, the correlation of data by statistical methods and, finally, interpretation of results. This indicates that this approach will be applicable in all the phases of this research, such as selection of text types to be included in the corpus, analysis tools and the interpretation of corpus evidence.

According to Laviosa (2002) corpus linguistics is not a linguistic theory but a methodology that can be applied to a wide range of linguistic enquiries. This study involves the application of corpus linguistic approaches in specialised bilingual lexicography. The reasons are well articulated by Teurbet (2001:127) in arguing that ‘lexicography is the second major field where corpus linguistics not only introduced

new methods, but also extended the entire scope of research, however, without putting too much emphasis on the theoretical aspects'. The combination of corpus linguistic and lexicographic theories is of great value for the present study. It is consequently very important to outline the lexicographical theories that are relevant to this study and also to show how they complement each other.

The other theories that are relevant for this study are Wiegand's general theory of lexicography and the theory of lexicographic functions. The latter theory has been developed by researchers from the Centre for Lexicography at Aarhus School of Business since the early 1990's. The main feature of this theory is that it is user-oriented (cf. Bergenholtz and Tarp, 1995, 2003; Tarp, 2000, 2008; Fuertes-Olivera, 2009). The theory takes into account the characteristics of specific users, the typical user situations, the typical problems users experience in such situations, their needs to address such problems and finally, the lexicographic functions of the available and prospective dictionaries regarding the assistance users may find to solve their problems (Nkomo, 2010:372). The two theories emphasise the determination of the functions and purpose of a dictionary. They are based on a functional approach and regard dictionaries as utility products that should be designed according to the needs of the users. On the basis of the said theories, the envisaged specialised dictionaries will assist the users in situations where they either want to acquire knowledge on specialised subject fields, or when they want to make use of LSP language for reading, writing or translating. The user-situations emphasised by the function theory, are cognitive and communication-oriented. This is extremely relevant in the case of isiXhosa because lexicographic resources are needed by the majority of users when there is a problem with comprehending texts with LSP vocabulary. The specialised texts, on the other hand, are the sources of information sharing; hence they need to be accessible to all official languages.

1.6 Data collection tools and procedures

The researcher collected and analysed a variety of texts from different sources, organisations or institutions and these were used to develop the English-isiXhosa Parallel Corpus (EngXhPC. These include:

- The Constitution of the Republic of South Africa and other Acts
- Annual Reports from government departments
- Annual Financial Statements from Provincial Treasury, Western Cape
- Information brochures from Financial Services Board of South Africa
- Bona Magazine articles on financial matters
- Pan South African Language Board (PanSalb) Annual Financial Statements.

The texts comprise written electronic texts in English as the source language and isiXhosa as the target language. The electronic format made the encoding process much faster and easier. These parallel texts, for example included annual reports, extracts from financial statements, budget speeches that are written and translated from English into isiXhosa and translated information brochures from other institutions. The following websites: www.gov.za,www.pansalb.org,https://www.fsb.co.za, and www.westerncape.gov.za were identified to select and retrieve these documents as they are posted for public consumption.

The permission to download and use these texts was sought from the owners beforehand and this was granted. The researcher selected these texts because they are authentic documents that are published by recognised financial institutions, academic institutions and government departments. The translations are often done by professional translators though it is important to note that lack of terminology and resources in African languages generally impacts on the quality of translations into these languages. “Texts that are produced by authors with proven credentials are likely to contain more authentic examples of LSP use than texts by authors who are not proven experts’ (Bowker and Pearson, 2002:54). In other words, authorship has an influence on the reliability and authenticity of the whole research. The credentials of the parallel texts included in the parallel corpus played a role in the selection.

1.6.1 Data analysis and interpretation

Corpus linguistics is firmly rooted in empirical and inductive analysis of data. A corpus cannot be manipulated without the use of computer software. In the proposed study, a corpus analysis tool will be used to analyse the EngXhPC and, as noted, the selected software is ParaConc: a parallel concordancer that was designed by Michael Barlow in 1995. The concordancer possesses the following features (Barlow 2003:1-4):

- Ability to sort and count words in a variety of ways
- Alignment of parallel (translated) texts
- Finds and displays in an easy-to-read format, in context all occurrences of a particular search term (and minor variations thereof)
- Identification of translation equivalents
- Highlighting potential translations
- A collocation viewer, which allows users to see which words belong together
- Frequency lists, etc.

By using Paraconc lexicographers can identify source terms and their equivalent translations, words in contexts, possible translations that could act as synonyms, and the most common translations for headword selection through the frequency list. As much as data analysis will be handled by ParaConc, the scrutiny and interpretation of corpus evidence will be done by the researcher. The features of this concordance that are relevant for this study are, a frequency list that generates an alphabetic list as well as ordering items according to frequency, bilingual concordances which display the texts in a Key Word in Context (KWIC) format, a collocation viewer etc.

1.6.2 Ethical considerations

Research ethics are focused on what is morally proper and improper when engaged with participants or when accessing information/data. During data collection, translated texts were collected to create the EngXhPC. These written texts are the intellectual property of either the author or the organisation that published them; therefore seeking permission for use is an ethical prerequisite. Although some parallel texts were available on relevant websites, the researcher wrote request letters to the various institutions to apply for permission and explain the purpose of the study. This approach

is in line with Atkins and Rundell (2008:82)'s assertion that: 'In order to ensure that your corpus is "in good legal health", you will first need to find out who owns the copyright of each text that you plan to include (this isn't always as straightforward as it sounds)'; In the request letters, for instance, the following points to the copyright owners were included:

- Texts will only be used for academic purposes
- The only information required for the corpus is financial terms and the context in which they are used.

In compliance with UNISA policy on research ethics, an application for ethical clearance was submitted to the Research Review committee before undertaking the research. This was done in order to ensure that the 'research is conducted with scholarly integrity, excellence, social responsibility and ethical behaviour'. The researcher further submitted a letter of request and consent to the office-bearers of the Department of Arts and Culture and Treasury in the Western Cape, Financial Services Board of South Africa, Bona Magazine stating the purpose of the research. The translated texts are available on their websites in one source language only, English; therefore, the translated copies in the target language required a request and permission letter. The ethical clearance certificate issued by the Research Review Committee, the letter of request and permission are attached as appendices.

1.7 Defining the key terms and concepts

Some terms and concepts carry different meanings. The key terms that are used frequently in this study are defined in relation to the subject of the study; these are corpus, parallel corpus, specialised corpus, ParaConc, English-isiXhosa Parallel Corpus (EngXhPC), a dictionary and specialised bilingual dictionary. The term corpus is first defined.

1.7.1 Defining a ‘corpus’

The definition of this concept is evolving with the times due to the introduction of language technology methods. In simpler terms a corpus is a Latin word that denotes a body of texts which has a plural form ‘corpora’. ‘A corpus can be described as a large collection of authentic texts that have been gathered in electronic form according to a specific set of criteria’ (Bowker and Pearson, 2002:9).

In addition, McEnery and Wilson (1996:87) define a corpus as ‘a body of texts which is carefully sampled to be maximally representative of a language or a language variety’. Finally, Baker (1995:225) points out that within corpus linguistics, generally, the term ‘corpus’ is usually used to mean ‘any collection of running texts held in electronic form and analysable automatically or semi-automatically (rather than manually)’. In a thorough analysis of the definitions offered above, one would agree that different scholars emphasise different aspects of a corpus. Bowker and Pearson (2002) note four important features: i) authentic, ii) electronic, iii), large, and iv) specific criteria. This means that a corpus is electronic if it is held and manipulated by means of a computer. These scholars further explain that an authentic text is an example of real, live language that consists of a genuine communication between people going about their normal business. Authenticity has become a focus of corpora design for dictionary writing because it replaces the intuitive based lexicography. This is relevant for this study, because real, live language is what the users of specialised dictionaries are looking for.

1.7.2 Parallel corpus

A parallel corpus can be simply defined as the source language and target language (ST-TT) pairs which are aligned. According to Atkins and Rundell (2008) the term denotes a set of corpora in which texts in language A correspond in some way to those in language B. A parallel corpus is composed of original texts and their translation; hence it is also called a translation corpus. Because of their efficiency, parallel corpora can be used by different user groups who are involved in different investigations. Bowker and Pearson (2002) mention three groups of people interested in using a parallel corpus: language teachers/learners, students and teachers of translators and computational linguists. Although the user group’s identification by Bowker and

Pearson is restricted and leaves out other users of a parallel corpus such as lexicographers, the amount of information that a parallel corpus offers to the lexicographer can improve the quality of bilingual dictionaries proposed in the proposed study.

1.7.3 A specialised corpus

A specialised corpus on the other hand comprises representative oral and or written texts which reflect the kind of language of a particular domain (Flowerdew, 2004:21). In addition, Bowker and Pearson (2002:12) define it ‘as the one that focuses on a particular aspect of language’. Bergenholtz and Tarp (1995) concur with these authors when affirming that a specialised corpus contains texts of a certain type, and aims to be representative of the language of this type. The nature of this resource implies a conscious selection of texts and computer software to extract the specific and specialised data information. The definition of corpus software used to construct and analyse the targeted corpus follows.

1.7.4 What is ParaConc?

‘ParaConc is a simple software program that makes it easy to analyse translated texts’ (Barlow, 2008:12). It is a search tool or corpus analysis tool that is designed for parallel texts and can handle up to four languages. It is also described as a multilingual concordancer. This software is selected for this study because of its application for parallel corpora and its ability to generate frequency lists, identify translation units, search terms in contexts and provide frequency counts. Section 2.5.5 and Chapter 3 provide more details about ParaConc.

1.7.5 Defining a dictionary

A dictionary according to Merriam-Webster's online dictionary is defined as follows:

1: a reference source in print or electronic form containing words usually alphabetically arranged along with information about their forms, pronunciations, functions, etymologies, meanings, and syntactical and idiomatic uses

2: a reference book listing alphabetically terms or names important to a particular subject or activity along with discussion of their meanings and applications

3: a reference book listing alphabetically the words of one language and showing their meanings or translations in another language

4: a computerised list (as of items of data or words) used as reference (as for information retrieval or word processing).

This definition is comprehensive and covers at least 3 types of dictionaries, namely, a general-purpose monolingual dictionary, a bilingual dictionary and a specialised monolingual dictionary. Nonetheless, this definition excludes any specialised bilingual dictionary. Of importance in the definition is the description of the type of information that is included in a dictionary or looked up for by the users.

A dictionary can be regarded as 'a text conveying information' (Svensén, 1993:2). The dictionary users consult a dictionary when they experience a knowledge gap, which is why the arrangement in alphabetic order is emphasised in the definitions. As the conveyor of information, a dictionary is expected to contain information that caters for different users with various needs. Having defined what a dictionary is, in the following section the two terms: 'bilingual and specialised' are explained.

1.7.6 Specialised bilingual dictionary

The focus of the proposed study is to show how bilingual information can be extracted from parallel corpora; therefore, it is imperative to first explain what specialised lexicography is: the sub-discipline of lexicography that deals with subject specific domains and covers language for special purposes. According to Bergenholtz and Tarp (1995:28) 'specialised lexicography is that branch of lexicography which is practised by LSP lexicographers who prepare specialised dictionaries'. The emphasis

of this explanation is that specialised dictionaries are the end products of the specialised lexicographic process. Specialised bilingual dictionaries according to Bowker (2010:157) are dictionaries that treat specialised fields of knowledge (e.g. business, chemistry, law). The main subject of these dictionaries is the language for special purposes (LSP) which consists of lexical items that are used to describe concepts in specific subject fields, such as technology, mathematics, business economics, finance etcetera. Due to their subject matter, specialised dictionaries can also be referred to as LSP dictionaries. In the current study, the term specialised dictionaries will be used more frequently in order to emphasise the type of terminology involved in the design of such dictionaries. The specialised texts of financial terms will involve English as the source language (SL) and isiXhosa as the target language (TL).

1.7.7 English-isiXhosa Parallel Corpus

An English-isiXhosa Parallel Corpus is designed by the researcher in order to fulfil the objectives of the current study. This EngXhPC will be restricted to the LSP specific to finance (cf. Bowker and Pearson, 2002). This type of corpus comprises written authentic texts in English and their translations in isiXhosa, meaning that English is used in the source text (ST) while isiXhosa is employed in the target text (TT). In terms of directionality of the corpus, EngXhPC is classified as mono-directional (English-isiXhosa) with a single subject-field of finance. The choice for this direction is based on the fact that most financial documents are translated from English. In addition, constructing a multi-field specialised corpus is a complex exercise that requires sub-corpora; hence a single-field parallel corpus is a selected choice for this study.

1.8 Chapter outline

The proposed study will be organised in the following manner:

Chapter 1 introduces the study by presenting the following information:

- Background and rationale
- Statement of the problem
- Aims of the study
- Methodology
- Ethical considerations.

Chapter 2 reviews literature that is relevant to the study. The review will emphasise the views of leading scholars on various themes that are important to the study. It will also indicate who has contributed to the field of corpus studies and dictionary making, as well as how their works are applicable to this study, the gaps that exist in this field and the contribution that will be made by this study.

Chapter 3 outlines the research methods and data collection procedures that will be used to achieve the above stated aims. The study will utilise a mixed method approach that combines the qualitative and quantitative aspects of research. An English-isiXhosa parallel corpus will be created and analysed through ParaConc to extract data that will be used to create a specialised financial English-isiXhosa bilingual dictionary.

Chapter 4 presents the findings and interprets them in line with the aims of the study.

Chapter 5 concludes the study by summarising the main points raised in the other chapters, pointing to the contribution of the study, identifying limitations and making suggestions for future research. This is followed by a list of references and appendices**.**

CHAPTER 2

Literature Review and Theoretical Framework

2.1 Introduction

This chapter presents a systematic review of literature that is relevant to the current study and an overview of theoretical frameworks that underpin this research. In order to situate the study, the researcher will first discuss corpus-based research conducted in Africa, with particular focus on the South African context. The theoretical overview will cover the development of translation studies and lexicography respectively.

2.1.1 Overview of corpus-based translation research in Southern Africa

This section presents a review of previous works related to the current study with a view to explaining what has been done in the field, the gaps, the relevance of selected literature and the reason why this study is being conducted. Corpus-based research related to the study has been conducted by various scholars from different angles, with the development of indigenous languages and terminology as the ultimate goal.

The use of corpora in developed languages such as English is at an advanced stage due to the availability of rich, accessible resources such as term banks, a variety of dictionaries, glossaries and the like. In the African languages, however, the use of this resource is still in its infancy, although it has to be mentioned that of late notable work has been done by African scholars in a bid to develop both languages and resources. The following are some of the scholars who have carried out research in this field.

Moropa (2005; 2007) designed a parallel corpus of English-Xhosa and used ParaConc to query the parallel corpus. Moropa (2005) investigated and analysed translation strategies used in the translation of official documents from English into isiXhosa. She also established whether corpus-based research could contribute to technical terminology development and translation strategies for translating technical texts. Through corpus-based research, it was observed that translation strategies commonly used by translators of the technical texts were simplification and explicitation.

The translation strategies enabled the translators to create appropriate technical terms. Although Moropa (2005) used a parallel corpus of technical texts, her focus was on translation strategies and term creation, whilst the present study investigates the use of parallel corpora of finance terms to extract bilingual terminology for bilingual lexicographic purposes. Moropa's study is relevant for the proposed study because it deals with financial terms in English and isiXhosa, the same language combination in the present study. In addition, Moropa makes a comparison between printed bilingual dictionaries of English to isiXhosa and information from the parallel corpora. The results proved that a carefully designed corpus is an invaluable resource for specialised terminology and a necessary aid for the translators. Moropa (2011) also investigated a link between simplification and explicitation and how the morphological systems of isiXhosa influence the two translation strategies. Other scholars who contributed to corpus-based translation studies in African languages are discussed below.

Ndlovu (2009) investigated translation strategies used by translators in Zulu health texts, making use of ParaConc in order to extract and analyse data. The study also aimed at examining the effectiveness of these translated texts in conveying the message to the target reader. Ndlovu (2009) compiled an English-Zulu health corpus. Of significance to this study, ParaConc was complemented by other research instruments: self-administered questionnaires, focus groups and semi-structured interviews. In motivating this approach Ndlovu writes, 'The analysis of the corpus alone would leave out important aspects of health texts' (Ndlovu, 2009:191). Ndhlovu (2012) investigated translation strategies used by English-Ndebele translators in translating specialised terms and taboo words. Her study focused on the Ndebele language spoken in Zimbabwe. She used HIV/AIDS texts in creating the English-Ndebele corpus. In the study Ndhlovu gives a detailed account of the functionality of ParaConc, such as word frequency counts, lists and hot words, thus illustrating the kind of valuable information that can be automatically retrieved from the electronic corpus. In a study conducted recently Ndhlovu (2016) investigated the use of ParaConc in extracting bilingual terminology. The two studies affirm the role of parallel corpora but further demonstrate that the lack of terminology is not unique to South Africa but a genuine problem of all African languages. The findings of this study are significant for the

current study because they indicate that the methodology selected in the current study is feasible when investigating a parallel corpus.

Nokele (2015) uses corpus-based research to identify the conceptual metaphor in the Nelson Mandela's autobiography, *Long Walk to Freedom*. Nokele investigates the metaphor from the English source text and then explores how the conceptual metaphors were rendered in isiXhosa and isiZulu target texts. The English-isiXhosa-isiZulu corpus was analysed through Paraconc. A comparison of how the metaphors were translated into isiXhosa and isiZulu revealed similar translator's styles. Her study is an affirmation of how ParaConc can effectively study the multilingual corpora regardless of the genre under investigation. The scholars discussed so far have selected ParaConc as a corpus analysis tool to manipulate a variety of text types. Other translation scholars who opted for other corpus software are worth mentioning. Parallel corpora can be used for various applications employing different corpus analysis tools as shown below.

Madiba (2004) conducted a corpus-based study with the aim of discussing the role of bilingual parallel corpora in the development of African languages, with special reference to Tshivenda. In the compilation and analysis of an English-Venda Parallel Corpus, he selected Multiconcord. Madiba focused on bilingual corpora for specialised texts with the ultimate goal of developing Special Language Corpora for African Languages (SpeLCAL). Madiba's study is pertinent to the present study because he views parallel corpora as 'a strategic means to fast-track the development of modern terminologies, bilingual dictionaries and glossaries indigenous languages of South Africa' (Madiba, 2004:133). Multiconcord was used to demonstrate that parallel corpora can be used as resources to study translation equivalents.

Other corpus-based translation studies conducted in South Africa were on interpreting (Wallmach, 2004) and Bible translations (cf. Masubelele, 2004; Naudé, 2004). These studies demonstrate how different types of corpora and corpus software can be used to conduct various investigations into different languages of South Africa. Wallmach's research (2004) focused on interpreting. The researcher provided insight into the nature of interpreted texts when the interpreter is under pressure. In querying the Gauteng Provincial Legislature Corpus that she set up, Wallmach made use of WordSmith Tools and ParaConc. The two features of WordSmith used were the Word lister and the concord, whilst ParaConc was then used to extract the source language

(SL) and the target language (TL) segments from the corpus. Wallmach (2004) discovered that the use of loan words was an indicator of increased cognitive load for the interpreters.

Corpus–based translation research has also been applied to Bible translations in order to describe and explain what existing translations look like. Masubelele (2004) examined the role played by Bible translations in the growth and development of written Zulu. Her article aimed at illustrating that translation policies adopted by translators may result in different linguistic choices. Masubelele (2004), in using a corpus-based approach, identified some of the variations used in 1986 and 1959 translation of the book of Matthew and further demonstrated that linguistic differences are actually a reflection of the different translation policies accounted for within DTS. A monolingual translational corpus of the two texts was used while WordSmith Tools was utilised to analyse data. The researcher found that the translators of the two isiZulu translations differ in their use of lexical items. The translators of the 1986 translation have adopted the norms of the target culture whilst in the 1959 translation the norms of the source text and culture were preferred. Besides a corpus-based approach, this study has applied the DTS research framework which is germane to the present study.

Finally, Naude (2004) also evaluated the Bible translations, focusing on the manner in which the predominant feature of biblical Hebrew poetry, namely, parallelism, is represented in various Afrikaans Bible translations. The type of corpus selected is a monolingual sub-corpus of Afrikaans translations of the Psalms. In this article, Naude (2004) used WordSmith Tools to perform type-token relation analysis. The researcher established that the translational patterns exemplified by means of explicitation and simplification are applicable to Afrikaans Bible translations. The trend shown so far is that WordSmith Tools is the most popular corpus analysis tool to query monolingual corpora.

Gauton and De Schryver (2004) demonstrated how special purpose multilingual and parallel corpora can be utilised as translation tools in finding suitable term equivalents when translating texts into isiZulu. They used two case studies, one a multilingual corpus with HIV/AIDS and the other a parallel corpus of labour-related terms. The corpus analysis tools used were WordSmith Tools and ParaConc. Of relevance to the current study is the observation that ParaConc is quicker to identify translation equivalents than querying a parallel corpus by means of WordSmith Tools. This is a

sound motivation for why ParaConc is a better choice for the proposed study. Additionally, ParaConc functions more efficiently with aligned translated texts.

From a different angle Prinsloo and De Schryver (2000) demonstrated the use of electronic corpora as the basis for the compilation of African languages dictionaries. Their area of focus was both the microstructure and macrostructure of dictionaries. Prinsloo and De Schryver (2000:304) argue that 'the field of electronic corpus lexicography is not an esoteric branch of linguistics that can only be pursued by scholars of English language. Even if the linguistic field of electronic corpus lexicography was invented by lexicographers of English language, corpus aided and corpus-based dictionaries for the African Languages are a reality too'. This study investigated the use of corpora in English and Sepedi. The types of parallel corpora used were general- purpose while the language pairs under investigation were Chiluba-Nederlands and English-Sepedi. These scholars stressed the importance of applying corpus-based methods in dictionary making in African languages. The text types selected for their corpora were literary works, magazines and suchlike. The researcher of the current study selected financial parallel texts with English as the source text and their translations in isiXhosa; the current study heeds the calls described above and shows how translation documents can be used to create bilingual dictionaries. The current study is methodological in approach, as such, it is important to situate the Corpus-based Translation Studies methodology within the larger theoretical framework. The section gives an overview of translation theories.

2.2 Theoretical overview

This section deals with the theories that explain and support the research problem of the present study. In the previous section, mention was made of the different benefits of corpus-based methods in terminology development and dictionary making. In achieving the objective of this study, theories from translation studies and lexicography are discussed in order to foreground the current research. To explain the phenomena of corpus studies in translation studies, it is neccessary to give a brief history of theories that contributed to the advent of this theory. In this regard, an overview of how translation studies developed from prescriptive to descriptive theories is provided. Because this study creates a synergy between translation studies and lexicography,

lexicographic theories will also be explored. An exposition follows, on the development of translation studies from linguistic-based theories to target-oriented approaches.

2.2.1 Linguistics-based theories of translation

This section offers a brief overview of how translation theories developed from linguistics-based theories through the cultural turn up until the incorporation of technological advances. The linguistic era of translation studies developed in the mid-twentieth century. The theories during this era brought about a remarkable shift in the history of translation studies, which was better known as the golden age of linguistic equivalence (Cheung, 2013:3). The linguistic theories were preceded by philosophical translation theories whose focus was on the source text and its message. The philosophical theorists preferred literal translation, while this was rejected in the linguistic era. The most influential proponents of linguistics-based theories are Jakobson (1959/2004), Catford (1965), Newmark (1981), Nida (1964) and Koller (1979/1989). Commenting on equivalence during this time Munday (2012:81) says the 'dominant linguistic terms in the linguistic turn were meaning and equivalence'. According to the review of literature, equivalence is regarded as the central issue in translation, although this created a controversial debate amongst translation scholars. Some scholars were in favour of equivalence in the translation process whilst others totally rejected it; still others revised it as a necessary requirement in translation theory and practice. Following is an overview of the diverse views on the notion of equivalence.

In an attempt to interpret the nature of equivalence Roman Jakobson (1959) introduced the 'notion of equivalence in difference' and thereafter suggested three kinds of translation, namely, intralingual, interlingual and intersemiotic translation. Interlingual translation is very significant in the current study because it refers to the translation between two languages. Although Jakobson emphasised the analysis of the source text and meaningful message transference to the target text, he admits that languages are different from one another; therefore, there can be 'no full equivalence between code units'. Intralingual translation according to Jakobson (1959) involves substituting messages in one language not for separate code-units but for entire messages in some other language. For Jakobson, meaning and equivalence are linked to the interlingual

form of translation, the latter being the type that takes place between two languages. Jakobson's views are relevant in the assessment of translated texts and their equivalence information to be extracted from the parallel corpora.

Nida (1964); Nida and Taber (1969) were also preoccupied by a concept of equivalence, but differently from Jakobson. Nida was a bible translator and his theory emanated from his practical experience. Nida (1964) rejected the notion of literal and word-for-word translation and came up with two types of equivalence: formal and dynamic equivalence. Formal equivalence 'focuses on the message itself, in both form and content, unlike dynamic equivalence which is based upon the principle of equivalent effect' (Nida, 1964:159). Formal equivalence is oriented to the source text whilst dynamic equivalence attempts a target-oriented approach to translation. Dynamic equivalence aims at addressing the reading and understanding of the Bible. Nida's introduction of the concept of formal and dynamic equivalence, according to Munday (2012:68), 'was crucial in introducing a receptor-based (or reader based) orientation to translation theory'. Nida was interested in the semantic value of a translation. However, Nida's work was criticised due to its bias towards Bible translation, while some critics even doubted its scientific approach. Despite the criticisms against Nida's systematic linguistic approach to translation, scholars such as Newmark were influenced by Nida's views. In an attempt to improve Nida's theory on equivalence, Newmark (1981) moves away from the receptor-based theory of Nida and proposes semantic and communicative equivalence, the former being source oriented and the latter being target-oriented. Newmark's concepts are based on function theory and place emphasis on text types and users. By introducing semantic and communicative equivalence, he was trying to narrow the gap that always exists between the source language and the target language.

Although Newmark and Nida's notion of equivalence exhibit some similarities, they differ on the point of emphasis. According to Nida the message of a translation should be in line with the both the cultural and linguistic needs of the receiver.

'Naturalness is a key requirement for Nida' (Munday, 2012:67). The desired goal of dynamic equivalence is to ensure that the source language is close to the target language. Newmark, on the other hand, suggests that translators should employ different translation methods for different types of texts, and emphasises faithfulness

to the readers, author and the source text. Nida and Newmark's theories contribute a great deal in the current study as they both emphasise content over form.

Catford (1965) followed a more linguistics-based approach to translation and insisted that 'the theory of translation is essentially a theory of applied linguistics' (Catford 1965:19). His contribution to translation theory was the introduction of types of translation and translation shifts. Catford (1965:21) described 'equivalent' as the key term in his translation theory. In improving on Nida's approach to translation equivalence, Catford made a distinction between formal correspondence and textual equivalence. 'A textual equivalent is any TL form (text or portion of text) which is observed on a particular occasion…to be the equivalent of a given SL form; a formal correspondent, on the other hand is any TL category which can be said to occupy, as nearly as possible, the same place in economy of the TL as the SL category occupies in the SL' (Catford, 1965:27). The two distinctions are describing the equivalence relations that are observed during the translation process. Catford's translation theory and the relations in equivalence makes a contribution in this study because it puts emphasis on translators striving for the closest equivalent when translating SL to TL.

The notion of equivalence that dominated the linguistics-based theories was criticised by other translation scholars (cf. Snell-Hornby, 2006; Basnett and Lefevere, 1990; Nord, 1997), whilst other scholars proposed an equal treatment with non-equivalence. These criticisms and challenges resulted in the redefinition of equivalence and the recognition of non-equivalence between languages. Baker (2011) noted that the difficulties of translating might be caused by non-equivalence. She acknowledged that non-equivalence creates translation problems that require different strategies to be employed by the translators. Mona Baker's levels of equivalence from word level and beyond contributed considerably in dealing with non-equivalence. Her translation strategies are commonly used by various translators in solving this problem.

Despite all the various challenges and criticisms on the notion of equivalence, recent translation research and theorists regard equivalence as a necessary requirement in translation practice: amongst the prominent ones are Baker (2011); Kashgary (2011) as well as Saule and Aisulu (2014). The researcher of this study is of the opinion that equivalence relations and different types of equivalents are salient elements in translation and bilingual dictionaries. In striving to achieve naturalness in translation and functional translations, translators have to think about the types of equivalence

that are suitable for different texts and their audience. The redefinition of, and the significant role of, equivalence is appropriate for the current study. Xiabin (2005:19) as cited by Kashgary (2011) attests that 'equivalence will remain central to the practice of translation even if it is marginalised by translation studies and translation theorists'. As long as identification of translation equivalents is the task of a translator, the concept will not easily disappear in the act of translating. The functional and target-oriented approach to translation builds a firm foundation for the type of dictionary envisaged in this study and its prospective use.

2.2.2 Functionalist theories to a 'cultural turn'

The functionalist approaches to translation emerged in the 1970s and 1980s as a significant shift from a linguistics-based approach to a sociocultural concept of translation. The functionalists rejected the concept of equivalence which was central during the linguistics-based era (Reiss and Vermeer, 1984; Nord, 1991, 1997). Functionalism constitutes a number of theories that were developed in Germany and Finland and was, according to Schäffner and Wiesemann (2001), motivated by the needs both of practical translation activities and of translator training. The most important theory on which functionalism is based on the Skopostheorie developed by Hans Vermeer in 1978 (Nord, 1997:138). This theory is based on the notion of skopos, a Greek term meaning purpose. The guiding principle of a translation process in terms of the skopos theory is the purpose or skopos. The main argument of functionalist approaches according to Schäffner and Wiesemann (2001:14) 'is the view that texts are produced and received with a specific purpose, or function, in mind'. Therefore, the focus in translation moved from source text to the target text and the target audience. The views by Nord (1997) and Vermeer (2000) are paramount in this regard. Vermeer views the change as the dethronement of the source texts whilst Nord regards the source text as the point of departure of any translation process. Another theory that is characteristic of functionalism, and valuable for the current study, is 'translational action'. Nord (1997:141) describes such action as 'the generic term coined by Justa Holz-Manttari in 1981 and designed to cover all forms of intercultural transfer, including those which do not involve any source or target texts'. This means that translation is a complex action that encompasses other actions such as analysing the user's needs, decision-making by the translator and an adequate end-product that fulfils the purpose

and function of the target text. ‘Translational action focuses very much on producing a target text that is functionally communicative for the receiver’ (Munday, 2012:121). The translator as the expert in the process should understand the reason behind translation of the ST and the function of the TT. These theories are related to the current study because the translated texts that are selected for analysis play a communicative role that is target specific. Extracting bilingual information from such texts contributes to the dictionary information that is directed to the needs of potential users.

2.2.3 ‘Cultural turn’ in Translation Studies

Another development in translation studies took place in the 1980s when translation was influenced by cultural studies. This shift was termed the ‘cultural turn’. ‘The term cultural turn is used in translation studies for the move towards analysis of translation from a cultural studies angle’ (Munday, 2012:191). The ‘turn’ downplayed linguistic-based models of translation; its emphasis was placed more on the cultural aspect of translation. The two pioneers of the ‘turn’, Bassnett and Lefevere, initially held varying views on what translation is. Lefevere is recognised as one of the leading theoreticians in the field of literary translation. In his works, Lefevere viewed translation as a form of language transfer, but influenced by culture. Bassnett, on the other hand, views translation not only as a kind of pure lingual activity but also as a type of intra-culture and inter-culture (Jixing, 2013:111). She saw translation as an inter-cultural activity. The cultural turn was formally announced by Bassnett and Lefevere in 1990 in a collection of essays entitled *Translation, History and Culture*. In this publication the two theorists agree that translation is an intercultural activity and that culture and language are intertwined. The cultural approach’s main concern is to situate all types of translation in the culture of the receptor language. The cultural turn theorists regard translation as a product of a broader cultural context that encompasses a number of belief systems. The cultural approach to translation indicates a shift from linguistics to extra-linguistics factors that prevail in the environment of the translated text. Unlike in source-oriented theories, the translational unit in this approach is culture. The interaction between language and culture became central during this period. The theorists of the cultural turn applied this shift to the literary texts and, as time went on, some scholars extended it to non-literary texts such as specialist ones (cf. Stolze, 2009; Szal, 2014). The cultural approach has been seen to cut across the literature vs.

non-literature divide as it ‘implicitly embraces all kinds of translation (Snell-Horny, 1990:84).

‘Cultural embedding as a feature of texts in general is also valid in technical and scientific texts’ (Stolze, 2009:124). During the translation of technical or specialised texts the translator is not only dealing with terminology but also bridging the gap between the two cultures so that the translated text is accepted in the target culture. In manipulating the source texts and the translations, the researcher should understand the cultural approach. Szal (2014:73) argues that ‘cultural elements in texts are present at all the linguistic levels - ranging from specific concepts and word forms, single sentences and text structures to pragmatic context’. The field of finance as the subject of the current research is a new form of knowledge in indigenous languages; therefore, subject-specific data needs to be easily communicated to the target reader requires cultural adaptation. In creating new terms based on English for isiXhosa, adaptations should consider culture. Wealth, for instance in the English culture, is always associated with money whilst in the culture of amaXhosa, the kraal constitutes the riches of a man (Moropa, 2007). In the olden days there were no banks and money was kept *kufelemntwini*; hence treasurer is translated as *unongxowa*. The relevance of this approach is explicated during analysis in chapter 4 of this study.

From the above discussion, it can be concluded that the cultural turn brought new changes to translation studies. The redefinition of translation as rewriting and intercultural communication is what motivated the selection of the cultural turn as the means not only to analyse the products of translation but also to understand the cultural factors that influence the translation process. The concept of translation as rewriting also redefined the role of the translator. The translators are required to consider various factors beyond language when translating. This shift helps to redirect the translation scholar’s excessive attention on the text to the surrounding socio-political, economic and cultural forces that impact on and constitute the text. Any translation takes the form of rewriting and is performed under certain constraints and for a specific function. The translations of specialised texts in a multilingual society are not exceptional. It is worth mentioning that the cultural turn does share some features with functionalism and polysystem theory. The polysystem theory is one of the translation theories that focused on target language culture and translation norms.

According to Gentzler (2001:70) the two most important shifts in theoretical development in translation theory have been:

1. The shift from source-text oriented theories to target-text oriented theories
2. The shift to include cultural factors as well as in the translation training models. Those advocating functionalist approaches have been pioneers in both areas.

Both functionalism and the cultural turn view translation as a communicative act that is aimed at serving a purpose for the target culture. However, the functionalists attach importance to the function of the target text. ‘The function of a translation in the target culture is determinative in the decision as to which aspects of the source text should be transferred to the translation’ (Nord 1991:6). It is the responsibility of a translator to ensure that a translated text (TT) is intelligible to the target readership by paying attention to the function, text type and cultural aspects of both the source language and the target language.

In sum, the focus of functionalism and the ‘cultural turn’ on the target audience, and the role of the translated text in the receiving culture, form the backbone of the study. The translated texts that are produced following the norms, cultural aspects and the purpose are likely to contain information that can also be regarded as functionally adequate. The specialised texts that were selected to construct the EngXhPC are intended to transmit a message to the target audience, so that the cultural context in which the text was produced is vital. Bilingual dictionaries that provide equivalents based on intuition do not serve the needs of users in a multilingual country such as South Africa; consequently, translators should be the mediators of the two cultures. Translation should accordingly be understood as an intercultural activity. DTS as a methodology that informs this study is briefly discussed below.

2.2.4 Descriptive Translation Studies (DTS)

Descriptive Translation Studies, abbreviated as DTS, is a theoretical framework that was proposed by Holmes and further developed by Toury (1995). Holmes (1972/2000) played a significant role in the development of translation studies. In an attempt to find a home for translation, Holmes (1972) made a survey of this field. In his popular work, 'The Name and Nature of Translation Studies', he proposed the name of the discipline as translation studies (TS) and its structure (cf. Toury, 1995; Snell-Hornby, 2006; Munday, 2012). Although there have been varying views about the scope of translation studies, translation saw a transformation and evolved over time. According to Munday (2012), one of the problems that led to these developments was the lack of an independent home for translation research which was dispersed across other older disciplines such as linguistics. Holmes's map of translation studies divides translation studies into two main divisions, namely, pure and applied translation studies. Pure translation studies are subdivided into theoretical and descriptive translation studies. Applied translation studies deal with translator training, translation aids - which concern lexicographical and terminological aids and grammar, translation policy and translation criticisms. All these developments gave birth to an independent field called DTS. Holmes' framework not only shows what constitutes translation studies but also demonstrates the interconnection of the various branches of the new field. DTS focus on three research areas:

- Product-oriented
- Function-oriented
- Process-oriented.

Product-oriented research examines existing translation by undertaking an analysis of source text and the target texts. Function oriented research on the other hand focuses on the function of the translated text in the recipient's socio-cultural environment.

This part of research puts emphasis on context rather than on a text. Process-oriented DTS is concerned with the decision-making process of the translator during translation.

Toury (1995) made further developments in the field of translation studies. He focused his attention on finding a methodology for DTS. Toury developed a descriptive and target-oriented approach to translation due to the influence of both Holmes and Even-

Zohar's polysystem theory. The polysystem theory places the translated texts in the historical and literary systems of the target culture. Munday (2012:169) explains that Toury's work calls for the development of a properly systematic descriptive branch of the discipline to replace isolated free-standing studies that are commonplace.

The key features of DTS are its being descriptive, target oriented, functional and systematic (cf. Snell-Hornby, 2006; Naude, 2005; Munday, 2012). These four features are fundamental in the current study, especially when they are tied up with Toury's claims that 'translations are the facts of target systems only'. 'The function of the translation in the target culture was the central premise of the new paradigm' (Snell-Hornby, 2006:49). The descriptive approach to translation studies is looking at the existing texts and describing the norms the texts seem to follow. In the current study, the terms extracted by ParaConc provide a picture of rules that may be followed during the translation process. The translations of financial texts are the mirrors of the target culture and the end users. The observation of empirical data in the form of translated texts is the object of this study. The applied translation studies branch in Holmes's map is also valuable in the current research because it taps into the bilingual dictionaries as translation aids. The dialectical nature between DTS and other two branches of translation studies, that is, theoretical and applied translations studies, strengthens the current empirical study which observes data for the purpose of dictionary-compilation.

From the above discussion, the tenets that inform DTS draw closer to corpus-based translation studies, CTS. This is a new theoretical framework that is influenced by both corpus linguistics and DTS. Laviosa (2002:16) presents the commonalities of DTS and CTS as follows:

1) Both embrace an empirical perspective and investigate their respective objects of study through the direct observation of real-life examples
2) Both approaches affirm that the generalisations derived from empirical evidence can only be valid if based on the study of large collections of texts, not just on individual instances
3) Finally, the principles that pertain to their respective objects of study are discovered by systematic research and are expressed in terms of probabilistic rules of behaviour rather than prescriptive pronouncements.

The two approaches emphasise descriptive methods and negate the prescriptive theories of the linguistic-based theories. The real-life examples evidenced from the translated texts represent a shift away from intuition because intuitive-based information is subjective and not reliable. The following section provides an overview of CTS, its links with descriptive translation studies and the impact on corpus design.

2.3 Corpus-based translation studies

This section presents information on CTS, which is the main guiding theory in this study, and corpus linguistics as the research paradigms that have transformed translation studies and other applied studies. According to Kruger (2002:70) corpus-based research emerged in the late 1990s. It is informed by a specific area of linguistics which involves the analysis of large corpora of authentic running texts, by means of computer software. The role of computers is significant in CTS as the texts are held in the electronic form. Although corpus-based research began in the 1990's in translation studies it remained neglected and was only strongly advocated by Baker in 1993. She laid down the agenda for corpus-based TS (1993 and 1995) and began collecting corpora of translated texts with a view to uncovering the distinctive patterns of translation (Laviosa, 2002). The interest grew out of the different needs of translation scholars, such as translation education, comprehension and terminology extraction (Sager, 1990; Laviosa, 2002). Subsequent to the development of CTS many corpora were being used for various investigations: English-French (Salkie, 1995), Churh and Gale in 1991 who used Canadian Hansard parliamentary proceedings; English-Italian (Marina et al 1992); English Norwegian (Johanson and Hofland, 1994; 1996). The subject of investigation was the translation product and the translation processes which are driven by DTS. The use of comparable and parallel corpora in the field of translation grew dramatically amongst language pairs.

Olohan and Baker (2000) proposed a Translational English Corpus (TEC): they explored the use of the parallel corpus option and found that better results were more frequent than in the comparable corpus. The Dutch Parallel Corpus (DPC), a 10 million item corpus which is described as a high quality, sentence aligned one was compiled in 2009 at the University College of Ghent in Belgium. This background demonstrates that CTS has advanced in the international world. The application of

multingual corpora according to Melby (2012) has moved beyond translation studies to machine translation (MT) and terminology management, while translation memory features in Translation Environment Tools (TEnTS) are also utilising multingual corpora. Google Translate is an example of machine translation that is based on multilingual corpora.

The utilisation of comparable and parallel corpora in South Africa has been sketched in Section 2.1.1 above. As indicated, corpora in Africa and South Africa in particular, have been developed by researchers for specific research questions. The applications of corpora are increasing in linguistics, interpreting translation, terminology management and lexicography. The various Universities in South Africa are also involved in corpus development projects. The University of Pretoria was the forunner in the compilation of corpora involving African languages. The North-West University's Autshumato is a Translation Environment Tool (TEnT) that is widely used by translators through the support of Department of Art and Culture (DAC). Corpus linguistic methods therefore are making tangible changes in linguistic investigations. Below is the definition of corpus linguistics and its characteristics. This information is relevant for the current research because it affords insight as to how this paradigm has revolutionised translation studies and dictionary compilation. Additionally, CTS is founded on corpus linguistics.

2.3.1 The definitions of corpus linguistics and characteristics

Corpus linguistics is described as the branch of linguistics that studies language by means of corpora. Notably, there are varying views on the scope and character of corpus linguistics; hence there are different definitions. These were mentioned earlier. Corpus linguistics is described as a methodology that is based on an empirical approach (Baker, 2010; Laviosa, 2004; McEnery and Wilson 1998; Teurbet, 2001).

These scholars further demonstrate various characteristics that fully recognise corpus linguistics as an empirical methodology. For instance, Biber, Conrad and Reppen (1998) describe corpus linguistics as having four main features:

- It is an empirical (experiment-based) approach in which patterns of language use that are observed in real texts (spoken and written) are analysed
- it uses a representative sample of the target language stored as an electronic database (corpus) as the basis of analysis
- it relies on computer software to count linguistic patterns as part of analysis
- and it depends on both qualitative and quantitative analytical techniques to interpret the findings.

Additional information on the scope of corpus linguistics is given by Teurbet (2001:129) who asserts that corpus linguistics extends our knowledge of language by combining three different approaches: i) the procedural identification of language data by analysis, ii) the correlation of data by statistical method, and finally, iii) the intellectual interpretation of results. These arguments emphasise the role of the computer during the analysis stage, but also recognise the significance of human judgement and the active participation of a researcher during the analysis stage. Corpus linguistics and its complex nature is a relevant methodology that can be successfully employed in the observation and analysis of a specialised corpus. Its applicability to a wide range of linguistic enquiries is the reason why it has been selected for this study. The meaning in its context is of growing interest amongst linguists. Teurbet (2004a:97) regards the study of meaning as the main concern of corpus linguistics, 'because corpus is a record of language as a social act (rather than a psychological phenomenon), and meaning too, is a social phenomenon'. The lack of contextual meaning and usage examples in the current bilingual dictionaries is a mounting concern; thus, the analysis of the naturally occurring texts in this study will provide adequate information required by the target users in the communicative and cognitive situations of specialised texts.

It is worth noting that within corpus linguistics, a distinction has been made between corpus-driven and corpus-based approaches. Tognin-Boneli (2001) draws this distinction, arguing that 'the main difference between corpus-based and corpus-driven approach is that the former starts with a pre-existing theory which is validated, while the latter builds up the theory step by step in the presence of the evidence, the observation of certain patterns lead to a hypothesis, which in turn leads to generalisations in terms of rules of usage and finally finds unification in a theoretical statement'. In support of Tognin-Boneli, Baker (2010:95) argues that corpus-driven

linguists tend to use a corpus in an inductive way in order to form hypotheses about language, and not making reference to existing linguistic frameworks'. Baker (2010:95) further contends that 'corpus-based linguists tend to use corpora in order to test or refine existing hypotheses taken from other sources'. According to Moropa (2007), one of the strengths of the corpus-based approach is that it can be applied to empirical investigations in almost any area of linguistics. In recent years one has also witnessed its application to other disciplines such as translation and lexicography. The corpus-based approach therefore remains a relevant framework in the design of a specialised bilingual dictionary of English and isiXhosa.

Olohan (2004:16) identifies the benefits of corpus-based methods in Translation Studies as being:

- an interest in the descriptive study of translations as they exist
- an interest in language as it is used in the translation product
- an interest in uncovering what is probable and typical in translation, and through this, in interpreting what is unusual
- a combining of quantitative and qualitative corpus-based analysis in the description, which may focus on a combination of lexis, syntax and discussion features
- application of the methodology to different types of translation, i.e. translations in different sociocultural settings, modes, etcetera.

The above points make a contribution in the current research because in addition to the ability to use large quantities of data on a computer, Olohan's emphasis on description, the language use in typical translation and the combination of qualitative and quantitative analysis provide reasons for the selection of this methodology.

Modern lexicography as a discipline that makes use of computers has benefited from the applications of corpus linguistics. The use of corpora in lexicography has dramatically changed the work of the lexicographer and brought about new methods in dictionary compilation. The benefits and effects of electronic corpora on lexicography have been acknowledged by a number of lexicographers and linguists (Krishnamurthy, 2006; Atkins and Rundell, 2008; Prinsloo and De Schryver, 2000). Rundell (2012:16) points out that corpus linguistics and the insights it has provided have broadened the scope of what lexicographers do. A list of these benefits is also outlined by

Krishnamurthy (2012) as: i) the relative frequency of meanings in corpora led to their reordering in dictionaries, often overturning the traditional placing of etymological meanings first, and ii) the abundance of corpus examples has amended the need to invent examples. With regard to African languages, corpora have also effected enormous changes, although this resource is still underutilised. Prinsloo (2009) states that a corpus is generally acknowledged as an indispensable resource for the creation of dictionaries and lexicographic tools and is a source for a variety of programs. In conclusion, corpora have introduced the following elements: objectivity, authenticity and reliability in content. Bowker and Pearson (2002) who devoted their work to specialised languages and guidelines on how to use corpora pointed to a number of advantages over other types of resources:

- Their electronic form means that corpora can be larger and more up-to-date than printed resources, and they can be searched more easily
- Corpora consists of authentic texts that can be used to find out what people do and do not say, as well as how often they say it
- Corpora can be used to conduct new investigations or to test an existing hypothesis
- Corpora can be fun and interesting to explore.

The previous section focused on the use of corpora in general with particular reference to dictionary compilation. It is now appropriate to briefly unpack and define what a corpus is.

2.3.2 The definition of corpus

To reiterate: the current study is corpus-based where an English-isiXhosa parallel corpus will be analysed using translation software. As such, it is important to define and explain what a corpus is, parallel corpus and translation software. A brief definition of corpus was given in section 1.7.1 of chapter 1. In this section, a more comprehensive definition is provided, looking at how corpus is defined by various theorists. As such, it is important to define and explain what a corpus is, as well as a parallel corpus and translation software. The word ‘corpus’ simply means ‘a collection of texts put together according to some informed criteria’ (Zanettin, 2012:7). A corpus has large quantities

of running texts which are held on a computer. The definitions of a corpus by different scholars show similarities and differences and some of these are discussed below.

Many scholars identified certain features and characteristics of this resource (Bowker and Pearson, 2002; Atkins and Rundell, 1998; McEnery and Wilson, 2004; McEnery, Xiao and Tono, 2006). The striking observation is the variation on the emphasis of these features, and this is evident in the definition offered by each scholar. McEnery and Wilson have put emphasis on sampling and representativeness. However, Bowker and Pearson (2002) note four important criteria: i) authentic, ii) electronic, iii) large and, iv) specific criteria. This means that a corpus is electronic in form and manipulated by means of computer software. Bowker and Pearson (2002:9) explain further, “if a text is authentic, that means it is an example of real live language and consists of a genuine communication between people going about their normal business’. Authenticity has become a focus of corpora designed for dictionary writing because it replaces intuitive based lexicography. A corpus in the present study means “any collection of running texts held in electronic form and analysable automatically or semi-automatically (rather than manually)’. The above definition is appropriately selected for this study because the financial parallel texts collected to compile the English-isiXhosa Parallel Corpus were electronically and ParaConc performs a semi-automatic analysis. It is crucial to briefly discuss the types of corpora and classifications and motivate for the type of corpus selected for the current study. In this current study, a corpus is designed and manipulated as a resource for the compilation of a bilingual dictionary of English to isiXhosa financial terms.

2.3.3 Corpus typology

The classification of corpora is another interesting and complex area in corpus linguistics. Corpora are invariably designed for different reasons and purposes; hence the diversity of typology (cf. McEnery and Wilson, 1998; Baker, 1995; Kenny, 2001; Fernandes, 2006; amongst others). The types of corpus that are relevant in translation studies may represent different languages and text types. A corpus may cover one or more languages. A corpus that contains texts in one language is known as a monolingual or reference corpus. Kenny (2001) further divides monolingual corpora into single translational and non-translational. In this view single translational corpora

consist of texts which are translated into one language, whilst non-translational corpora contain texts that are original texts in one language. The British National Corpus (BNC) is the most familiar monolingual corpus that comprises English original texts which are not translations. Monolingual corpora may be either general or specialized. According to Zanettin (2011:15), 'general language corpora are created with the aim of representing as far as possible a given (national variety of a) language, and thus to function as a reference for that language'. Specialised corpora are restricted and focus on a specific variety of language.

Monolingual corpora in recent times have not only been used to solve general linguistic problems, but also play a crucial role in language learning, translator training and retrieval of information. Monolingual translational corpora are used by scholars who are interested in studying translated texts and translator's styles (cf. Kenny 2001; Bowker 1998). A discussion follows, of Baker's typology which is more relevant to corpus-based translation.

Baker (1995) classifies electronic corpora into three types, namely, multilingual, comparable and parallel corpora, whilst Fernandes (2006) in his revisiting of Baker's typology argues for only two types of corpora, comparable and parallel. Fernandes argues that the term multilingual does not have any contrastive features that could make it distinctive from other types. Fernandes (2006) regards his classification as more flexible than Baker's typology. Baker presents her three types of corpus as follows:

2.3.3.1 Multilingual corpora

Baker (1995:232) describes a multilingual corpus as 'sets of two or more monolingual corpora in different languages built up in either the same or different institutions on the basis of similar design criteria'. This type of corpus constitutes texts that are not translated. Such a corpus according to Aijmer and Altenberg (1996) can be used for contrastive linguistic work including bilingual lexicography. Although Baker (1995) proposes a multilingual corpus she expresses reservations about its usefulness in both theoretical translation studies and contrastive studies. The non-alignment in a multilingual corpus makes it difficult to match the information. Multilingual corpora may

be comparable or parallel. It is therefore, important to clearly differentiate a comparable corpus from a parallel one. The definition of comparable corpora is provided below.

2.3.3.2 Comparable corpora

A comparable corpus according to Baker (1995) consists of a collection of texts originally written in a language, say English, alongside a collection of texts translated into English. The two texts in terms of this definition share the same language and constitute translations. However, Alternberg and Granger's (2002) comparable corpora consist of original texts in each language, matched as far as possible in terms of text types, subject matter and communicative function. The comparison or similarity of the texts is based on a variety of criteria. In the classification of translation-oriented corpora, Zanettin (2012:11) distinguishes between comparable and parallel corpora:

> Comparable corpora can be monolingual, bilingual or multilingual, and are composed of texts which have no direct translational relation, i.e, they are not translations of each other. In contrast, parallel corpora (bilingual or multilingual) have a translation relationship. This defining relationship can be either unidirectional, going from one source text to a target language or bidirectional, going both ways.

In contrast to Baker's definition, Teurbet (2002) and Zanettin (2012) echo that the texts in the comparable corpora are not translations of each other. In addition, Zanettin's definition incorporates the number of languages involved as well as directionality. The texts in a comparable corpus should display similarities; nevertheless, there is no consensus on the nature of similarity, as is confirmed in the above studies. From the above different definitions, though, one can note that comparable corpora are useful in studies where scholars are contrasting or comparing similar texts. According to Kenny (2001), this type of corpus is central to research into translation universals. A discussion of the parallel corpora follows.

2.3.3.3 Parallel corpora

Parallel corpora are collections of aligned source language texts and target language texts (cf. section 1.7.2) The texts are translations of each other. Teurbet (2002:204) describes them as 'repositories of translation units and their equivalents'. They are very useful in translation studies and bilingual lexicography. The alignment technique allows for a valuable correspondence of the text in both languages. 'Parallel corpora can be unidirectional i.e. source texts in language A and target texts in language B, or bidirectional, i.e. source texts in language A and their translations in language A' (Olohan, 2004:24). A parallel corpus that consists of one language and its translation is termed *bilingual* whereas one with source text with translations in more than two languages is *multilingual.* A parallel corpus can further be divided into a general-purpose or special-purpose corpus. This type of corpus is restricted and focussed. A *specialised corpus* on the other hand comprises representative oral and or written texts which reflect the kind of language of a particular domain (Flowerdew, 2004:21). Bowker and Pearson (2002:12) add that 'a special purpose corpus is one that focuses on a particular aspect of language'. Bergenholtz and Tarp (1995) earlier pointed out that a specialised corpus contains texts of a certain type and aims to be representative of the language of this type. The nature of this resource implies a conscious selection of texts and the use of computer software to extract the specific and specialised data information. Both parallel and comparable corpora are commonly used in translation studies and bilingual lexicography. A specialised parallel corpus was developed in the current study with the aim of investigating translated texts of financial terms, which comprises a specialised domain.

The user group identified by Bowker and Pearson (see section 1.7.2) is restricted and leaves out other users of a parallel corpus such as lexicographers. For the purpose of this study, though, lexicographers are included in the list. In the following section, the use of parallel corpora is discussed in order to demonstrate its contribution to both translation and lexicography.

2.4 The role of parallel corpora in translation and bilingual lexicography

Parallel corpora according to McEnery and Xiao (2008) are a good basis for studying how an idea in one language is conveyed in another language. In translation studies parallel corpora have demonstrated usefulness as terminology resources for both professional translators and trainee translators. Aijmer and Alternberg (1996:12) regard parallel corpora useful in translation studies because they:

1) Give insight into the languages under study
2) Are ideal for comparison purposes
3) Highlight the differences between the ST and TT
4) Are useful for practical application in language teaching, translation studies, lexicography, etcetera.

The value of parallel corpora in bilingual lexicography is attested by a number of linguists and lexicographers. The make-up of these corpora and the number of languages involved makes it possible to compile bilingual dictionaries particularly for term extraction. Because a parallel corpus consists of a source text and its translation, this makes it possible to extract bilingual terminology that could be useful in dictionary writing. Teurbet (2007:53) argues that 'parallel corpora are especially helpful in the case of context-dependent translation equivalents' This is made feasible by the fact that aligned texts display sentence by sentence, which also allows one even to understand phrases and multiword expressions. McEnery and Xiao (2008) believe that 'the amount of information that a parallel corpus offers to lexicographers can considerably improve the quality of bilingual dictionaries'. There is general consensus that dictionaries compiled without the use of corpora lack the lexicographic information that is desirable to the users.

The use of parallel corpora in South Africa can improve the status of previously disadvantaged languages by increasing scarce terminology in the specialised fields. The terms that are created and extracted from the corpora contribute to standardisation. The increased use of these resources will also help in designing the specialised resources that can be applied in various linguistic investigations. Parallel corpora cannot be useful unless interrogated by means of suitable corpus analysis

tools. It is therefore important to explain what corpus analysis tools are and present an overview of analysis tools available in the market, and thereafter discuss in detail the tool that will be used in the current study.

2.5 Corpus analysis software

The number of corpus tools available has grown over the past thirty years, as not only lexicographers but also scholars from other linguistic sub-disciplines have become aware of the potential of corpora (Kilgarriff and Kosem, 2012:32). As such, corpora have become an invaluable resource for many language practitioners. These corpora nevertheless, cannot be manipulated without a variety of sophisticated corpus software. Baker (2010:102) explains that a stand-alone corpus is not particularly useful in terms of aiding linguistic enquiry, and notes that corpora are normally used in conjunction with software, which is able to carry out counting, sorting, and presentation of language features. Kilgarriff and Rundell (2006:121) add that a corpus is of optimal use to lexicographers if it is loaded into a corpus query tool which supports them in finding, collocational and grammatical patterns. In consequence, the resource may be useless without carefully selected corpus tools which are able to process and display the kind of information and results that are required by the researcher and lexicographer.

Pastor (2007:3) defines a concordancer as ‘a software tool that queries a corpus in order to locate and display each instance of a given node and the context in which it occurs’. This software is not tied to a particular language and can also be used by translators, linguists, teachers, lexicographers and others interested in the analysis of multilingual texts (http://www.athel.com/parapaper.html accessed on 05/11/2016). Research has proven that corpus software or tools are designed and used according to the different types of corpora and projects at hand. Corpus analysis tools are diverse and are able to perform a variety of functionalities (cf. Kilgarriff and Kosem, 2012; Bowker and Pearson, 2002; Moropa, 2005; De Schryver and Prinsloo, 2004). Hunston (2006:234) provides an outline of what corpus software does:

- Searches the corpus for a given target item
- Counts the number of instances of the target item in the corpus and calculates relative frequencies
- Displays instances of the target item so that corpus user can carry out further investigations.

The technological and linguistic advances in the English language have made it possible for many analysis tools to be used to study the language. These include: ParaConc (2001), AntConc (2004), WordSmith Tools (1996), SketchEngine (2004), MonoConc, Pro (1995) among others. Antconc is free while the rest are commercial products. The access to a corpus software is very important when one makes a selection. A number of corpus tools are for local installation which means that one uploads these on one's computer and uses them elsewhere. The web-based corpus tools are office bound and are only operational at a fixed place. ParaConc is a simple accessible software that can be installed in a personal computer.

In navigating and analysing the EngXhPC that will be created in this study, the researcher will use the parallel concordance software known as ParaConc. Barlow (2003) describes ParaConc as a tool designed for linguists and translators who wish to work with translated texts. ParaConc is a well-known bilingual or multilingual concordancer that is designed to work with parallel or translated texts. ParaConc can accept up to four languages, a source text and three target texts or languages. The functions of ParaConc include concordances, collocation tables, word frequency lists, regular expression search as well as search options and hot word utilities. The features that are significant in the current study are: ParaConc's ability to create word lists (alphabetic and frequency order), provide statistical information, sort functions, identify hot words, and generate concordances. The success of these features depends on the alignment of the parallel texts. The next section presents an overview of corpus design processes.

2.6 Corpus design

Designing a corpus is a decision-making process regardless of the type of corpora and project at hand. Atkins and Rundell (2008:57) elaborate that designing a corpus means making decisions about:

- How large it will be
- Which broad categories of text it will include
- What proportions of each category it will include
- Which individual texts it will include.

These decisions are always taken at the planning stage and guide the researcher towards the fundamental features of the corpus to be designed. There seems to be a consensus amongst researchers on the factors that should be considered when designing a corpus. These include size, balance, representativeness, copyright and permissions. However, the debate continues as to whether some of the factors are attainable in actual practice. Biber (1993) and Bowker and Pearson (2002), for example, regard representativeness to be an essential consideration, whilst Atkins and Rundell (2008) argue against the focus on representativity. In Atkins and Rundell's words, 'a truly representative corpus is an impossible goal because we are sampling from a population whose nature is unknowable and whose extent is unlimited' (Atkins and Rundell, 2008:75). This argument may be relevant for a general language corpus because it is not easy to represent the whole language for general purposes; nevertheless, in a special purpose corpus it can be achieved.

Another factor in corpus design that is also debatable is the size. Is a large corpus a better one? Various scholars offer different answers to this question. Atkins and Rundell (2008) advocate for a large corpus in order to obtain adequate information on rarer words and rarer usages.

Their argument is really based on the general language dictionaries. Bowker and Pearson (2002) focus on specialised language and assert that it is important not to assume that bigger is better. However, they acknowledge that one may get more useful information from a small corpus that is well designed. From these arguments it is clear that designing a corpus depends on the nature and the purpose of the project as well as the potential users. In creating a specialised bilingual dictionary, a large

representative parallel corpus is required, but for purposes of this study that wants to show how bilingual terminology can be extracted, a smaller representative corpus will suffice. The previous section provided an overview of how translation studies evolved. Because a parallel corpus will be used to extract bilingual information that can be used in bilingual dictionary making, it is necessary to furnish a brief overview of lexicographic issues such as the development of lexicography from traditional methods to corpus-based lexicography, lexicographic theories that are relevant to the present study, and relations with translation studies.

2.7 The development of lexicography

The making of dictionaries resides in lexicography. In fulfilling the purpose of the current study, the latter term is explained in this section. The development of lexicography into a scientific discipline that is informed by recent theories and methods means that it makes a major contribution to this study. Before discussing bilingual lexicography as a sub-discipline of lexicography, an overview of what it is in general will be beneficial for the development of my argument. Lexicography has been defined and described from different angles by linguists and lexicographers due to various reasons, such as the progression of time and different schools of thought. For a long time, lexicography has been understood as the art and craft of dictionary writing.

According to Svensén (1993:1), 'lexicography is a branch of applied linguistics which consists in observing, collecting, selecting, and describing units from the stock of words and word combinations in one or more languages'. Svensén (1993) regards it as a branch of applied linguistics, whilst Wiegand (1984) defines lexicography as a scientific field concerned with the production of reference works on language. Both scholars agree on the practical side of lexicography yet differ on the independence of this practical activity. Hartmann and James (1998) on the other hand define lexicography as:

> The professional activity and academic field concerned with dictionaries and other reference works. It has two basic divisions: lexicographic practice, or dictionary-making, and lexicographic theory, or dictionary research.

Hartmann and James (1998) also emphasise the independence of lexicography as a field and further expand a little, to incorporate the theoretical side of this scientific discipline. In the recent literature reviewed, there is a consensus on the nature of lexicography as an independent field which is divided into practical and theoretical lexicography. According to Gouws and Prinsloo (2005), the practical component of lexicography developed well into the second half of the twentieth century before it was complemented by a theoretical component. For a long time in the history of lexicography, practical lexicography has been at the centre; hence, as noted, many definitions focus on dictionary-writing. The present study follows the 'two-fold' character of lexicography. The compilation of modern dictionaries should consider both sides of the coin. The theoretical component focuses on research regarding, e.g., the form, contents and functions of dictionaries (Gouws and Prinsloo, 2005:1). In exploring the use of corpora in the compilation of English-isiXhosa bilingual specialised dictionaries, the researcher's ultimate goal is to ensure the improvement of quality and contents of these dictionaries in order to serve the needs of the users.

The other element that is prevalent in the concept of lexicography is whether it is a sub-discipline of linguistics or an independent discipline. In terms of pure linguistics studies, lexicography has been regarded as a sub-discipline of linguistics (Zgusta, 1971). In current research on modern lexicography, however, some scholars regard lexicography as an independent discipline (Tarp, 2000; Bergenholtz and Gouws, 2012).

The approach followed in this study is an integrated approach that defines lexicography as an independent discipline which is concerned with the practice and theory of dictionaries in co-operation with other theories. The positioning of lexicography as a pure linguistic discipline is not supported by the researcher of the current study (cf. Zgusta, 1971). As an independent field of study, lexicography has its own theories and principles. In practice, dictionary writing requires the application of sound linguistic principles. According to Zgusta (1971:15), 'lexicography is a very difficult sphere of linguistic activity'. Zgusta further positions theoretical lexicography within the broader linguistic framework. In his contribution it is evident that linguistic developments have a significant influence on the changes observed, both practically and theoretically. However, the general consensus amongst scholars in 21st century is the emphasis on the independence of lexicography and the influence of linguistics. The positioning of

lexicography within a broader linguistic framework benefits the current study because producing high quality specialised dictionaries depends on both lexicographic and linguistic theories which are sound.

Gouws and Prinsloo (2005:5) concludes, 'Although, lexicography is not regarded as a sub-discipline of linguistics, the strong link between linguistics and lexicography, both theoretical lexicography and the practice of the compilation of special language dictionaries, may never be ignored'. Because a dictionary deals with the recording and processing of linguistic signs, a linguistic background is unavoidable. The independence of a field does not mean isolation. The changes and improvements are sometimes informed by its interaction with other related fields. The current study affirms the interface of the two disciplines, because good dictionaries can be produced when principles of linguistics are taken into consideration. In any attempt by lexicographers to improve dictionaries and their contents, a thorough investigation of new developments in linguistics can be beneficial in all the stages of dictionary making. As indicated, the ultimate goal of practical lexicography is a dictionary and the information that is extracted is used in the design of a bilingual specialised dictionary. The next section will give a brief overview of different types of dictionaries, classification models and approaches prevalent in modern lexicography, dictionary uses and users, and finally state the reasons why a specialised bilingual dictionary is valuable in the current research.

2.7.1 Different types of dictionaries

The writing and design of dictionaries is as old as written language; nevertheless, the classification of dictionaries is still a complex debate. Dictionaries available in both English and isiXhosa differ considerably in different ways, such as size, content, language involved and mode. The reason behind this is due to the various methods and approaches in classifying them. Dictionaries, therefore, can be classified into various types on the basis of different criteria. In doing so some scholars have followed the typological classification based on several criteria, parameters and distinctive features (cf. Al Kasimi, 1971; Singh, 1982; Svensén, 1993); while others have opted for a broader categorisation of dictionaries (cf. Zgusta, 1971; Bergenholtz and Tarp, 1995; Gouws and Prinsloo, 2005).

Zgusta (1971) provides a broader classification of dictionaries into two major divisions: linguistic and non-linguistic dictionaries. A linguistic dictionary is concerned with the words of that language. Examples of non-linguistic dictionaries are encyclopaedias. According to Gouws and Prinsloo (2005:48), encyclopaedic dictionaries are directed at the extra-linguistic features of the items to be treated, whereas linguistic dictionaries focus on the linguistic and pragmatic aspects such as word categories and usage examples respectively.

Having discussed the various approaches in the classification of dictionaries by various prominent lexicographers, the researcher in this study is going to describe the basic types of dictionaries into monolingual, bilingual, general purpose dictionaries and specialised dictionaries. The classification selected for this study is an attempt to illustrate the need for specialised dictionaries in isiXhosa which necessitated this study despite the other types of dictionaries available in isiXhosa.

2.7.1.1 General dictionary vs. specialised dictionary

Dictionaries can be classified as general or special. The distinction by Al-Kasimi (1971:31) regards 'a general dictionary as one which attempts the coverage of the whole lexicon of the language whereas a special dictionary deals with one sector of the lexicon.' In contrast to the technical or specialised dictionary, a general dictionary concentrates on the general vocabulary (LGP) of the language. General dictionaries aim at recording the whole language, including specialist terms to a certain extent. In lexicographic practice, a general dictionary cannot omit the technical terms that are used in daily language as a result of technological advancement (cf. Svensén, 1993; Gouws and Prinsloo, 2005). Such dictionaries include a wide variety of information categories and therefore assist the general users to obtain various types of information. Specialised dictionaries are referred to as restricted because they deal with language for special purposes and record terminology of a special field; hence they are called LSP or specialised dictionaries. They might be related to a specific semantic field or specific subject domain, like a dictionary of synonyms or accounting.

2.7.1.2 Monolingual and bilingual dictionary

This classification is based on the number of languages involved in a dictionary: 'The monolingual dictionary describes a language by means of that language itself: it gives the meanings of words by means of definitions or explanatory paraphrases' (Svensén, 1993:20). It involves one language and is often meant for the speakers of that specific language. Monolingual dictionaries can be subdivided into comprehensive, standard, desk and pedagogical dictionaries. A comprehensive dictionary represents large and voluminous dictionaries that describe the standard language, e.g. the *Oxford English Dictionary* (*OED*). Both the macro-structure and micro-structure are comprehensive. Such dictionaries are comprehensive, record a number of information categories and can be used to look up information such as definitions, pronunciation, example sentences, collocation, idioms and the like. A standard or desk dictionary is distinct from a comprehensive one. Although they also describe the standard language, they are smaller than a comprehensive dictionary because they deal with a certain part of the lexicon. *Isichazi-magama sesiXhosa* (ISX) (2008) is a typical example of a standard monolingual dictionary.

Unlike a monolingual dictionary, 'A bilingual dictionary uses two languages, one as the object of description and another as the instrument of description' (Stark, 2011:13). Thus it involves two languages, the source and the target language.

According to Gouws and Prinsloo (2005:151), 'bilingual dictionaries can be regarded as one of the typological categories most frequently used by the average member of speech communities in a multilingual environment'. Such bilingual dictionaries may serve either a decoding or an encoding function or both. A decoding dictionary can be used for reading, writing and comprehension. Most often, specialised bilingual dictionaries are used for the decoding function, which is why they are regarded as amongst the best translator's tools. The function of bilingual dictionaries and the user's needs determine the type of lexicographic information to be included. The distinct feature of bilingual dictionaries is a translation equivalent. Because of different user situations, bilingual dictionaries include a variety of data categories. An ideal entry of a bilingual dictionary according to Zgusta (1971) should include grammatical, pronunciation, semantic and pragmatic information. In South Africa, the number of bilingual dictionaries surpasses monolingual dictionaries because of their utilisation by a variety of users (Mongwe, 2006; Mafela 2005). In isiXhosa, in particular, 90% of

existing dictionaries are either bilingual or multilingual (see section 2.7.1 above). Bilingual dictionaries are used as the instruments for different lexicographic needs such as language learning, cognition, documentation and standardisation (cf. Gouws, 1996; Nkomo and Wababa, 2013; Klein, 2010). Bilingual dictionaries can further be divided into mono-directional and bi-directional varieties. This study will focus only on a specialised bilingual dictionary. A discussion of specialised bilingual dictionaries, the type of bilingual dictionary specific to this study, follows.

2.7.2 Specialised bilingual lexicography

Lexicography is always classified as general or specialised lexicography. According to Bergenholtz & Tarp (1995:28) 'specialised lexicography is that branch of lexicography which is practised by LSP lexicographers who prepare specialised dictionaries'. The emphasis of this explanation is that specialised dictionaries are the end products of the specialised lexicographic processes. The main subject of these dictionaries is the language for special purposes (LSP) which consists of lexical items that are used to describe concepts in specific subject fields, such as technology, mathematics, business economics, finance and so forth. Bowker and Pearson (2002) describe LSP in opposition to general purpose language (LGP). They define LGP as the language that we use every day to talk about ordinary things in a variety of common situations. In contrast, LSP is the language that is used to discuss specialised fields of knowledge. The practice of lexicography has focussed on general-purpose lexicography for many decades, which resulted in scanty research on specialised lexicography. However, specialised lexicographers started showing a growing interest in this field recently (Bergenholtz and Tarp, 1995; Marza, 2009; Nielsen, 2014). These studies focus on the need to improve research and practice and on the use of the specialised dictionaries as specialist tools. The work by Bergenholtz and Tarp (1995) in the *Manual for Specialised Lexicography* makes a considerable contribution in this study. Bergenholtz and Tarp (1995:11) sum up the concern of LSP lexicography: As a special part of lexicography in general, LSP lexicography certainly does work with LSP terms:

- LSP lexicography works with both systematic and alphabetic macro-structures, deciding in each case which is appropriate
- LSP lexicography must necessarily to a greater extent be both descriptive and prescriptive
- Addresses itself to laypeople and experts alike
- LSP lexicography prepares dictionaries for both encoding and decoding.

In addition to the above functions, Fuertes-Olivera and Nielsen (2012) describe specialised lexicography as a branch of lexicography that analyses, describes, designs and produces specialised information tools to satisfy user's timely needs in a cognitive, communicative or operative usage situation. Fuertes-Olivera agrees with Bergenholtz and Tarp (1995) on the two functions of specialised dictionaries but adds an operative function. The two functions that are relevant for this study are the communicative and cognitive functions.

As already mentioned, for many decades specialised lexicography, in particular bilingual lexicography, did not receive as much attention as general monolingual lexicography. This resulted in few bilingual specialised dictionaries, offering inadequate information, in both English and other languages. In English, however, the number of specialised bilingual dictionaries is commendable whilst the situation in African languages is lamentable.

The challenges noted above do affect African languages more. The status of bilingual lexicography generally and specialised lexicography in particular needs to be addressed in response to new theories and technological advancements. Moropa (2005) highlights the inadequacies found in the current bilingual dictionaries of English –Xhosa, and therefore recommends the use of electronic parallel corpora as resources for translators. This clearly shows that currently, bilingual dictionaries cannot serve a variety of needs of dictionary users who are in specialised fields. The types of information tools that are required should be multi-functional and serve the diverse needs of users.

The current study is responding to the various studies conducted with regard to a need to produce specialised dictionaries in the indigenous languages of South Africa (Klein, 2009; Moropa, 2005; Chabata, 2013; Nkomo, 2010, amongst others). These authors agree that African languages need to be developed by utilising them in all the domains

of life and addressing the lack of specialised terminologies. In responding to these genuine calls, this study seeks to close the gap by exploring a corpus-based approach that will be able to produce dictionaries that will not only improve the quality of specialised dictionaries but also be target-oriented.

In the African context, also, very few experiences of specialised lexicography exist (Ndinga-Koumba-Binza 2011; Chabata 2013; Nkomo 2010). The neglect of bilingual or multilingual specialised lexicography in the indigenous languages should belong to history. The position taken in this study is to improve that situation, focussing also on the role specialised bilingual dictionaries are playing in a diverse multilingual country such as South Africa.

Chabata (2013) proposes the compilation of specialised dictionaries in indigenous languages and explains why. He sees the compilation of specialised dictionaries as a step in raising the status of indigenous African languages in preparation for their use in high profile functions in society. Chabata (2013:58) further argues that 'specialised dictionaries are a sure way of empowering the languages so that they become usable in all domains of life, including those that are highly technical'. In South Africa, there is a growing demand for translating specialised texts from English into various African languages. These efforts cannot be achieved without adequate specialised bilingual dictionaries. The role of such dictionaries in indigenous African languages is further affirmed by Nkomo (2010). According to Nkomo (2010:386) 'LSP dictionaries in African languages may be useful for functions such as text production, text reception, translation and acquisition of specialised encyclopaedic and cultural knowledge'. The daily invasion by new concepts and terms requires either the compilation of new products or the updating of the existing ones. The relevance of specialised bilingual dictionaries in African languages during this era of globalisation should not be underestimated in South Africa.

The following section furnishes a brief background of how dictionary making moved from traditional methods to computerisation. This information assists in understanding the platform that was created by computers for electronic manipulation of data and the reason why the corpus method is selected in this study.

2.7.3 Dictionary compilation: from traditional methods to computerisation

Computer technology plays an important role in the compilation of both the monolingual and the bilingual dictionaries. Before the advent of computers, practical lexicography involved a long and frustrating process where lexicographers had to use manual methods that took a long time to produce a dictionary. The compilers had to collect citations from literature manually. In the case of indigenous languages where literature was limited, information was gathered by fieldwork and sometimes by using informants. The information would be transcribed and recorded in the slips or cards that were manually sorted in an alphabetic order. This method created many challenges that resulted in years and years of dictionary-making. Lew (2013) contends that the method was laborious in the extreme, and it also had a methodological flaw. The majority of bilingual dictionaries in English-isiXhosa were not compiled according to the principles of modern lexicography. All the bilingual and multilingual dictionaries of isiXhosa mentioned in chapter 1 of this study are the products of the traditional methods of lexicography. The compilation of the *GDX* (1989) Volume 3 (Q-Z) for example, took almost 20 years. The collection of sources, sorting and drafting of dictionary entries were all done manually by a team of lexicographers. The compilation of specialised dictionaries without the use of a computer results in inadequate products that compel the users to resort to other inappropriate lexicographic products. Research conducted in both lexicography and linguistics has attributed the shift from a manual approach, into the computer-based approach, to modern linguistics. This study, therefore, seeks to fill that gap and ensure the improvement of the dictionary compilation process, resulting in quality products.

Studies that have been conducted dealt with the general-purpose monolingual dictionaries. Bergenholtz and Tarp (1995) focus on specialised lexicography and affirm that a computer is a necessary tool in specialised dictionary-making because:

- It is fast, both at retrieving material and performing operations on it
- It is consistent and capable of performing complicated operations on complicated material with greater precision
- It is not fatigued by the monotonous repetition of the same operation over and over again

- It is becoming faster and faster and more compact and at the same time its storage capacity is increasing, and finally
- The price: performance ratio is becoming increasingly favourable.

The focus of these two authors forms the basis of the aim of the current research. As noted, corpus based specialised dictionaries in African languages are few and sometimes non-existent. Designing a specialised bilingual dictionary is a complicated exercise that involves the treatment of a special language for a specific purpose. The researchers above not only emphasise the efficiency of computer technology but also the avoidance of methodological flaws during the design process. The manual approach was dependent on the intuition of the lexicographers and thus resulted in inconsistencies and omission of significant data information. A brief discussion of the lexicographic theories that underpin the present study follows.

2.8 Lexicographic theories

Section 2.3 demonstrated how lexicography and bilingual lexicography are viewed by different scholars. The different definitions attached to this scientific discipline are indicative of the various theories that exist. In order to position my study, a background on relevant theories is essential. A dictionary and the kind of information included in a dictionary is the outcome of the practical side of lexicography. The methods, principles and the content structure of dictionary are informed by theoretical lexicography. The insight of lexicographic theories provides a theoretical background on the characteristics, purpose and the functions of bilingual dictionaries. In a discussion of lexicographic theories, it is important to realise that there is not only one single theory but, rather, different theories (Gouws, 2012:456).

Although literature in lexicography reveals that a number of lexicographers and scholars do not believe that lexicography has a theory, let alone theories (cf. Bejoint, 2010; Atkins and Rundell, 2008), this study draws from those theories that are applied in specialised lexicography; whilst on the other hand a large number of recent studies offer evidence of lexicographic theories that have improved lexicography as an independent discipline (cf. Wiegand, 1984; Bergenholtz and Tarp, 1995). The two lexicographic theories that are relevant are Wiegand's general theory and the theory

of lexicographic functions. The following sub-sections will explain each theory, their tenets and relevance to the current study.

2.8.1 Wiegand's general theory of lexicography

Wiegand is regarded as the first lexicographer to describe lexicography as an independent discipline. His predecessors classified lexicography as a sub-branch of applied linguistics. Wiegand (1984:13) argues that lexicography is not a sub-branch of linguistics or lexicology and that 'it is more than the application of linguistic theories and methods or the utilisation of linguistic philological findings'. Wiegand's words motivate his general theory of lexicography which regards lexicography as independent discipline, and therefore moves away from the reliance on linguistics theories alone. Wiegand's general theory has a comprehensive structure that includes amongst other factors: the structure of dictionaries, genuine purpose, the relationships with other theories, history of lexicography, the organisation theory. The current study regards Wiegand's general theory as the basis of lexicography. The key elements that are selected from the said theory in designing specialised bilingual dictionaries from English to isiXhosa can be listed as follows:

- A dictionary as the 'utility tool' that serves the needs of the society
- The genuine purpose of dictionaries
- Relationships with other theories.

According to Wiegand (1984) the general purposes of various dictionary types are derived from the communicative and cognitive needs of the society. In consequence dictionaries are designed for various reasons; therefore, different dictionaries are compiled for different purposes. This research is an attempt to provide insight into the need for specialised dictionaries that are designed for specific purposes. Another element of the general theory of lexicography that intrigues the writer of the current study is the connection with other theories. In producing high quality specialised dictionaries that can serve a variety of uses, I am of the view that theories and strategies from translation studies are necessary. The application of theories and practice of other disciplines is motivated by a number of scholars who support

Wiegand's theory (Hartmann and James, 1998; Gouws, 2012; Rigual and Calvi, 2014). Gouws (2012:457) also reiterates:

> Although lexicography is an independent discipline it has links and mutual components with a number of disciplines and the theories of these disciplines can play a role that also qualify their contribution as being of a theoretical nature.

Wiegand's general theory of lexicography is still relevant in specialised lexicography of African languages and is discussed and applied by current lexicographers (cf. Smit, 1996; Gouws and Prinsloo, 2005; Nkomo, 2010). The relevance of this theory in the current study is attributed to its emphasis on the genuine purpose of the dictionary as the 'utility tool', and also the recognition of lexicography as an independent discipline which has links with other disciplines. This study is also grounded on the theory of DTS which stresses the interdisciplinary nature of the discipline of translation studies.

2.8.2 The theory of lexicographic functions

The modern theory of lexicographic functions was developed at the Centre for Lexicography by a group of scholars at Aarhus School of Business in the early 1990's. The publications on the modern theory of lexicographic functions are evident in the works of its founders, Bergenholtz and Tarp (1995, 2003; Fuertes-Olivera and Tarp 2014). It is described as the functional approach and called function theory or theory of lexicographic functions (Tarp, 2008). The two concepts are used interchangeably in this study. Fuertes-Olivera (2009:122) is of the view that dictionaries are utility tools that are made to satisfy a specific type of need that arises in a certain type of user within a certain type of situation. This theory is user-oriented; hence the users are at the centre of its methods. Hartmann (2004) asserts that the lexicographic theory first of all focuses on the potential dictionary user. This term refers to the target user of the dictionary. Tarp (2008:40) describes the cause of this approach as a 'shift from the actual dictionary users and dictionary usage to potential users and social situations in which they participate' There are four main elements that constitute the theory of lexicographic functions: i) the dictionary user, ii) user-situation, iii) user's need, and iv) lexicographic assistance. Therefore, in planning a dictionary as a utility tool, the intended user's need should be taken into consideration. Bergenholtz and Tarp (2003) sum this up:

> Lexicographers, therefore, have to make a profile of the intended user group and a typology of the user-situations where problems or needs may pop up that can be solved by providing lexicographic data. On this basis the functions and purpose of a dictionary can be determined.

The functional approach to lexicography according to Bergnholtz and Tarp considers lexicography as an independent discipline and dictionaries as utility products that are meant to satisfy the needs of the users. The identification of lexicographic functions influences the content and the form of a dictionary. The application of this theory in specialised lexicography is valuable for this study. Consequently, there are researchers who put emphasis on user groups and their specific needs (cf. Gouws & Prinsloo, 2005; Bergenholtz & Tarp, 1995). The two types of functions advocated by the function theory of lexicography are communicative and cognitive-oriented functions of dictionaries (Bergenholtz and Nielsen, 2006). The communication-oriented functions deal with problem-solving information such as writing and speaking, reading and listening and translation whereas the cognition-oriented function focuses on acquisition of information and knowledge.

In the past, specialised dictionaries were aimed at the experts of different specialist fields. Bergenholtz and Tarp 1995:51) explain that 'It is often desirable that specialised bilingual dictionaries fulfil a combination of different functions, meaning that they should not be limited to being, for instance, pure translation dictionaries.' The desire for the combination of functions implies the presentation of various types of information in a dictionary. The potential users of such dictionaries could be experts, semi-experts and interested laypersons.

The focus of this research falls on how best the two lexicographic theories can be applied in the compilation of specialised bilingual dictionaries of English and isiXhosa. Wiegand's theory is useful in this study because it deals with the structure of dictionaries and their role. The function theory puts more emphasis on dictionaries as utility tools that should be designed according to the needs of the users. The function theory strengthens the relationships with other theories from other sister disciplines for the production of high quality polyfunctional dictionaries. The kinds of specialised dictionaries proposed in this study should be target-oriented and serve various needs of different users such as teachers, researchers and translators.

2.8.3 Links between translation studies and bilingual lexicography

Translation studies is defined as a distinct discipline whose main objectives are: i) to describe the phenomenon of translating and translation and ii) to establish theories that can be used to explain the two facts (Munday, 2012). There are several meanings that can be attached to the concept of translation. It can be defined as:

(i) the general subject field or phenomenon
(ii) the product - that is, the text that is translated
(iii) the process of producing the translation (Munday, 2012:8).

Translation may denote a discipline, product (translated text) and or the act of translating. The link being discussed in this section is between lexicography and translation, that is, the phenomenon and the product. Research on the links between the two disciplines is very sparse, although 'the two fields according to Hartmann (2004:18) have a special responsibility for describing and reducing interlingual contrasts and thus helping more people cross language barrier'. The focus on what is common between the two disciplines could bridge the unnecessary gap which existed. However, a number of studies focus on the differences and interdisciplinarity of translation studies and lexicography, leaving the close relations between the two less addressed. Hartmann (2004) questions the gap that exists between translation and lexicography and calls for closer relations. As an experienced lexicographer, Hartmann observes close relations in both theory and practice. Rigual and Calvi (2014:39) also argue that the relationship between the two disciplines is not as smooth as one would wish. In their study Rigual and Calvi (2014) present similarities and differences and thereafter propose close relations between translation and lexicography due to the number of similarities. Hartmann (2004:11) provides a useful understanding of what translation can do for lexicography:

> Translation is relevant to lexicography in two ways: as supplier of translation equivalents to be included in the bilingual dictionary and a consumer of information made available by a lexicographer to professional translators.

Hartmann's argument above is biased and does not state the relevance of lexicography to translation studies. However, Humble (2010 in Rigual and Calvi 2014) argues that bilingual lexicographers are indeed translators; not only because they conduct translation tasks frequently, but also because compiling a bilingual dictionary

itself is a translation undertaking as it involves translating all the lexical items from one language to another.

These arguments strengthen the close relationship between bilingual lexicography and translation. Indeed, in compiling a bilingual dictionary a lexicographer needs translation competence that is based on sound translation theories. In my opinion, translation and lexicography should be more closely connected in a multilingual country like South Africa. The studies should place more emphasis on the relations than on differences. The effective communication among different language groups is impossible without translation and bilingual lexicography. The translation of specialised texts depends on sound application of translation theories; therefore, the two disciplines do influence one another. The links between translation and lexicography, with particular focus on bilingual lexicography, are advantageous for this study. The particular focus on the relations strengthens the interdisciplinary nature of the two disciplines and provides further answers on the selection of theories made by the current study. The previous section focused on lexicography in general, following is an overview of how isiXhosa lexicography developed.

2.9 The development of isiXhosa lexicography: An overview

Having discussed the development of lexicography as a scientific discipline, it is important to give a brief history of how isiXhosa lexicography has developed from 1889, when the first standard dictionary was published, to the present scenario. This background will help to situate the current study within the larger framework. The history of isiXhosa lexicography dates back as early as 1772. Merely a few studies have been conducted on the history of isiXhosa lexicography (cf. Pahl et al 1989; Mtuze, 1992; Moropa and Kruger, 2000; Nkomo and Wababa, 2013). The mentioned scholars conducted various studies in which the beginning of isiXhosa lexicography from the arrival of the missionaries in Southern Africa becomes evident. This is also captured in Mtuze's words when he writes, 'Xhosa lexicographical studies cannot make sense until we realise how it was possible for foreign scholars to get involved in this demanding and highly academic linguistic exercise given the fact that they were all non-mother-tongue speakers of the language' (Mtuze, 1992:165). In 1992, Mtuze conducted a critical survey on the development of dictionary making among the

amaXhosa. His work is not only a critical commentary on the earlier dictionaries but also an evaluation of *The Greater Dictionary of Xhosa* published at the University of Fort Hare. The lexicographic history of isiXhosa is presented in terms of (i) earlier history and (ii) the contemporary Xhosa 'lexicographical scenario'. Moropa and Kruger (2000) in investigating the 'Mistranslations of culture-specific terms in Kropf's Kafir-English Dictionary' provide a biographical sketch of Kropf (an outstanding lexicographer of that time) and a very detailed exposition of the missionary contribution to isiXhosa lexicography. Nkomo and Wababa (2013) present a critical overview of isiXhosa lexicography from the past, present and to the future. The available literature always makes reference to the missionary efforts in developing isiXhosa as a national language.

In the introductory part of *The Greater Dictionary of Xhosa*, the Editor-in-Chief, Professor Pahl, provides a rich history of Xhosa lexicography. The visit of Andrew Sparrman, the pioneer of lexicography, between 1772 to 1777 marked the beginning of practical lexicography in the history of amaXhosa (Pahl, 1989:xxxvi). Andrew Sparrman's contribution to the lexicography of isiXhosa emanates from his interest in the isiXhosa numerals, nouns, adjectives and verbs. Following the pioneering work by Sparrman, Pahl furnishes us with a sequence of major works that preceded the actual publication of the first dictionary of isiXhosa:

i) Barrow's short list of words with their Hottentot equivalents in the years 1797-1798
ii) *A vocabulary of the Kaffra Language* by Dr Vanderkemp published between 1795-1802
iii) In 1826, the Lovedale Press issued *A Systematic Vocabulary of the Kaffrarian Language* by John Bennie
iv) *A Vocabulary of the Kaffir Language* by John Ayliff was published from Wesleyan Mission House, London in 1846.

The above grammatical works are detailed accounts of preliminary work carried out by the missionaries. Although these attempts were aimed at learning isiXhosa for communicative reasons, the outputs resulted in the reduction of isiXhosa into writing, the standardised orthography and the development of the language in general (cf. Mtuze, 1992; Moropa and Kruger, 2000). In Mtuze (1992:2)'s words, 'The grammatical

work had to be established before any attempts at recording and describing language and its sound system could be made'. Since IsiXhosa was a spoken language with no recorded lexicon it would not have been possible to record the language without an acceptable orthography and list of words.

The arrival of Appleyard marked a significant progression in dictionary writing but the manuscript of his *Kaffir Dictionary* could not be published due to the 1850 wars (Pahl 1989: xxxviii). Dohne's *Zulu-Kaffir Dictionary Etymologically Explained* was published in 1857, in Cape Town. Dohne's publication was based on the previous lexicographic works. It was an advanced study that marked the beginning of a new phase in Xhosa lexicography. Other dictionaries that appeared during this phase were Davis's *Dictionary of the Kaffir Language* in 1872 together with *Xhosa and Zulu Dialects*. These were followed in 1877 by an *English and Kaffir Dictionary*. Both Dohne and Davis's works were so systematic that they built a strong foundation for Kropf's standard dictionary. The *Kaffir-English Dictionary* by Dr Albert Kropf was published in 1899 at Lovedale Mission Press in Alice. Although other dictionaries were compiled between 1872 and 1877, *the Kaffir-English* was accepted as the standard dictionary of isiXhosa as it superseded previous ones (Pahl et al 1989). Kropf's dictionary was later updated by Godfrey and published in 1915. 'One of the aims of the *Kaffir-English Dictionary* was to help non-speakers of isiXhosa, especially the missionaries of the time who wished to learn the language or to improve their knowledge' (Moropa and Kruger, 2000:72). Kropf's dictionary is still regarded as one of the most useful dictionaries amongst translators and other language practitioners.

According to Pahl (1989:xxxviii), a number of dictionaries appeared after 'Kropf's dictionary', some from commercial publishers; amongst these are:

i) *English-Xhosa Xhosa-English* Dictionary in 1950 and
ii) *Afrikaanse Xhosa Xhosa-Afrikaanse Woordeboek*, both published by Via Afrika Publishers in 1950
iii) *The Concise Trilingual Dictionary in English, Xhosa, Afrikaans* and *Die Kort Drietalige Woordeboek in Afrikaans, Xhosa and Engels* by Jennings were published by Lovedale Press in 1961
iv) In 1976 a school dictionary, *the Xhosa Dictionary: English-Xhosa-Afrikaans, Xhosa-English-Afrikaans* by Nabe et al was published by Thandapers

v) K B Hartshorne et al compiled the *Dictionary of Basic English-Xhosa Across The Curriculum* which was published in 1984

vi) In 1985, the *English-Xhosa Dictionary* by Arnold Fischer et al was published by Oxford University Press. Fischer's bilingual dictionary came at a time when Kropf's dictionary was out of print. Pahl (1989) describes it as a 'godsend'.

The establishment of the Xhosa Dictionary Project (XDP) at the University of Fort Hare in 1968 marked another advancement in isiXhosa lexicography. This initiative was informed by the need for a modern scientific standard dictionary experienced by the students of that time. The acute need was realised after 40 years when the *Greater Dictionary of Xhosa* (Volume 3 Q-Z, henceforth *GDX*) was published at the University of Fort Hare in 1989 by Prof Pahl as Editor-in-Chief. The *GDX* is the first comprehensive trilingual dictionary with isiXhosa as lemmatising and defining language. The planning and location of *GDX* was a culmination of the co-ordinated approach towards lexicographic practice. In 2002, the XDP was reconstituted into the isiXhosa National Lexicography Unit as per the mandate of the Pan South African Language Board (PanSALB) Act of 1995 (Act 59 of 1995) later amended in 1999.The two volumes of the GDX were published under the auspices of this Unit established by PanSALB at the University of Fort Hare. The *GDX* Volume 2 (K-J) was published in 2003, whilst Volume 1 (A-J) was published in 2006. The two volumes were the products of computerisation, unlike Volume 3. The use of computer software started in 1989 and fast tracked the lexicographic processes such as data capturing, editing and proofreading. There are notable improvements in the latest volumes of *GDX*; however, all three volumes of *GDX* contain a comprehensive lemmata that includes the following features: lemma, grammatical information, definitions, idiomatic expressions and usage examples. There are also rich addenda which cover linguistic, historical, anthropological and cultural aspects of isiXhosa.

Subsequent dictionaries published at the XNLU were *Isichazi-magama sesiXhosa (ISX)*, the first ever monolingual dictionary of isiXhosa by Tshabe (Editor-in Chief), Nokele and Guzana (2008) which was followed by *Isichazi-magama seMathematika neNzululwazi* (2014) by Wababa, a specialised dictionary targeting school learners from grade 1-7. The release of this dictionary is a major step towards the specialised lexicography of isiXhosa. It is the first of its kind from the XLNU. The dictionary entry

consists of English headwords, part of speech labels, equivalents or explanations with usage examples.

In sum, the dictionaries that were published previously are mostly bilingual and multilingual ones. This was due to pressing needs of learning the Xhosa language for communication purposes. In other words, the efforts of missionaries in the compilation of bilingual dictionaries were based on the urgent need to learn the languages of the communities in order to spread the word and to communicate. However, this scenario has not changed if one also looks at the current dictionaries. From 1989 to date the lexicography of isiXhosa has been dominated by various types of bilingual dictionaries but most of them are general-pupose bilingual dictionaries. This is attributed to the needs of the users and policy requirements. South Africa is a multilingual country; therefore bi/multilingual dictionaries will remain useful tools in learning languages and enhancing daily communication activities.

Methodologically, the information entered in these dictionaries was gathered from sources such as previous dictionaries, written literature and vocabulary/word lists. A shift from this traditional approach was brought about by the *Oxford IsiXhosa-English English-isiXhosa School Dictionary (2014).* In the introduction this dictionary is described as a 'new type of bilingual dictionary'. The collection method is well-captured by De Schryver et al (2014:xi) in the following explanation:

Headwords were selected for their high frequency in a large corpus of words in isiXhosa. A corpus is drawn from hundreds of texts containing millions of words that have been taken form a wide range of sources- novels, textbooks, official documents, and others.

The *Oxford isiXhosa-English School Dictionary* is the first corpus-based dictionary in the history of isiXhosa lexicography. Its corpus consists of both written and spoken texts. The frequency counts were used as selection criteria. The selection and sorting were performed electronically and words that appeared often were included, whilst rare words were omitted. The meanings of words and authentic examples are also extracted from the corpus.

The year 2014 may be regarded as one of great achievements in isiXhosa bilingual lexicography since the first specialised dictionary was published by XNLU and a corpus-based dictionary by Oxford University Press. Despite the fact that the two

dictionaries are school dictionaries, the two publications are significant for the current study because they illustrate the inadequate number of specialised and corpus-based dictionaries in isiXhosa. The current study fills a gap in the lexicography by suggesting the use of parallel corpora in addressing the long-felt need of specialised dictionaries for a variety of users. The use of parallel corpora and ParaConc in making blingual dictionaries will provide lexicographers with a quick and efficient method that provides terms, their translations, synonyms and contextual usage in one sweep, thereby eliminating the monotony of dictionary making.

It has been learnt from the preceding section that the dictionaries were used as standardisation tools looking at terminology development and orthography. This background on isiXhosa lexicography would not be complete without touching on critical issues that reflect the nature and the structure of the language. The following section sketches the development of isiXhosa, its morphological system, orthography and terminology development.

2.10 IsiXhosa morphological system

The aim of this section is to highlight the distinctive features of isiXhosa as the indigenous language that is studied together with English in the present study. English belongs to a group of languages where words are always written disjunctively. IsiXhosa is one of the indigenous languages of South Africa and the second most spoken official language in South Africa, following isiZulu (cf. 2011 census). IsiXhosa is a member of the Bantu languages group and an agglutinative language which is characterised by a common noun classification system. The different morphemes in agglutinative languages are glued together to form one word. The two features of the morphology that are surveyed are the noun class system and the concordial system. The noun class system and concordial systems are indispensable features of isiXhosa in corpora and in dictionary-making. In isiXhosa lexicography it is the practice to explain the grammar of the language in the front matter (see Pahl et al 1989; De Schryver and Reynolds, 2014) and preceding the analysis or interpretation of corpus evidence (cf. Moropa, 2005; Ndhlovu, 2012). The reason for the discussion of these two features is to provide insight into the analysis of EngXhPC and into the extraction of bilingual terminology suitable for dictionary making.

A number of studies on the morphology of African languages had been carried out. In isiXhosa the works of Pahl (1971), Satyo (1983) and others have made a massive contribution in understanding the grammar of isiXhosa. Like all other Bantu languages, isiXhosa is characterised by a rich morphology, noun class system with closely related concordial system agreement (Gxilishe, de Villiers and de Villiers, 2007; Du Plessis, 1997). The Meinhof noun class system is used in isiXhosa due to its international status (Satyo, 1983). The nouns in isiXhosa are classified into Classes 1-17, based on the prefixes they take. Prefixation according to Katamba (2003) is regarded as a hallmark of Bantu morphology. The morphological system of a language is very vital in corpora because words occur in context and the harmony that is created between the noun and other parts of speech in a sentence is useful for the analyses. The numerous noun classes of isiXhosa and prefixes are tabulated below.

Table 2.1: IsiXhosa noun classes, prefixes and subject concords (SC)

Class Number	Prefix	Noun	SL noun	Subject concord
Class 1	um-	***um****thengi*	buyer	u-
Class 1(a)	u-	***u****somashishini*	businessman	u-
Class 2	aba-, abe	***aba****thengi*	buyers	ba-
Class 2 (a)	oo-	***oo****somashishini*	businessmen	ba-
Class 3	um-	***um****vuzo*	salary	U
Class 4	Imi-	***imi****vuzo*	salaries	i-
Class 5	Ili-,	***I****tyala*	debt	li-
Class 6	Ama-	***ama****tyala*	debts	a-
Class 7	Isi-	*isibonelelo*	allowance	si-
Class 8	Izi-	***izi****bonelelo*	allowances	zi-
Class 9	i-, in-	***i****akhawunti,* ***in****tlawulo*	account, payment	i-
Class 10	ii-, iin, izin-	*liakhawunti, iintlawulo*	accounts, payments	ii-, zi-
Class11	ulu-, ulw-,	*uhlahlo-lwabiwo-mali*	budget	lu-
Class 14	ubu-	*ubutyebi*	wealth	bu-
Class 15	uku-, uk-	*ukuhlawula, ukonga*	pay, save	ku-

The table demonstrates the class nouns of isiXhosa according to Prof Carl Meinhof's classification with their class prefixes. The bolded formatives in the examples are typical class prefixes. Every noun belongs to a group of other nouns which are semantically related. Satyo (1983:25) explains the relation of class 1 and 2, 'Ubukhulu becala izibizo zeli hlelo zizibizo ezibhekiselele ebantwini. Le nto phofu ingathethi kuthi

kwamanye amahlelo azikho izibizo ezisingisa ebantwini' (Most often, the noun in this class refers to human beings). This simply means that every noun belongs to a specific group that corresponds with the noun's prefix and the other concords such as the subject concord, as in Table 2.1 above.

Moropa (2005:82) describes 'a concordial system as frequent repetition of certain morphemes in the same sentence, and this promotes the euphony of the language'. The concords in isiXhosa are derived from the noun prefixes. The class prefix of class 2 is *aba-* and the subject concord is *ba-* as shown in Table 2.1. Katamba (2003) states that the noun prefixes play a crucial role in the extensive system of concord, i.e. agreement in Bantu languages. IsiXhosa has two types of concords, namely, predicative and qualificative concords. The predicative concord is found in the predicate, a verb. It may be a subject concord or object concord. The last column in Table 2.1 only shows how subject concord is derived from each noun class. The subject concord is selected because in the sentence it is obligatory, unlike the object concord. The qualificative concords are found in qualificatives such as pronouns, adverbs, adjectives. Due to the scope of this study these will not be discussed in depth as there are numerous grammar books and studies in English and isiXhosa that have dealt already with this information. A simple sentence in isiXhosa contains a subject, a predicate and an object (Subject Verb Object SVO).

1. *Umthengi **u**hlawula ityala*. (A consumer pays the bill).
2. *Umthengi uya**li**hlawula ityala.*

The subject concord u-prefixied to the stem –hlawula relates to the subject noun umthengi. The information on the morphological system of isiXhosa as an agglutinative language is relevant for this study because it assists in explaining the syntactic patterns displayed by the words in the concordances. Understanding how various concords such as object, adjectival, adverbial, etcetera, are linked to the subject noun is fundamental in corpus analysis. The morphology of a language is important in establishing how words are written. Following is a brief overview of isiXhosa orthography and its development as conveyed in the translated texts.

2.10.1 IsiXhosa standard orthography

The growth of isiXhosa and massive translations that are done at various levels has brought in new words that sometimes introduced new speech sounds. Due to the linguistic dimensions, orthographical changes cannot be left unattended as they impact on the structure of a language (cf. Probert and De Vos, 2016). New dictionaries and other written sources have to comply with the standard orthography. In section 2.7.1 of this chapter, scholars gave a glimpse of how isiXhosa developed from spoken language to a written language, and of efforts to write grammars and term lists in order to standardise the language before it was recorded in dictionaries. Louwrens and Poulos (2006) emphasise that words are defined through the structure of a language and that words are mediated through orthographies. In order to reflect a developed and growing language the orthography should be standardised for uniformity. In an attempt to fully standardise the African languages, PanSALB in 2005 embarked on the revision and modernisation of all nine indigenous languages in South Africa. 'Orthographies, as systems of encoding the written forms of a language, are inextricably intertwined with issues of standardization' (De Vos, Van der Merwe and Van der Mescht, 2015:162). This explanation highlights the importance of orthography and its impilcations for lexicography. The more uniform the writing system, the further the standardisation has progressed. According to PanSALB (2008), the revision process of isiXhosa orthography entailed the revision of the booklet, *Xhosa Terminology and Orthography No. 3* which was printed in 1980 and the introduction of a number of new rules (PanSALB 2008). The changes made were due to gaps and inconsistencies that were experienced by language practitioners; hence a new section on editing matters was introduced.

The period between the 2005 spelling and orthography rules and 1980 indicates that some of the inconsistencies were already entrenched in the writing system. The nature of the written words and their syntactic behaviour is always mirrored in large texts such as corpora and dictionaries (Probert and De Vos, 2016). These resources reflect the status of the language structure; accordingly, standardisation is supposed to be a prompt and ongoing process. Information intended for inclusion in a dictionary has to be accurate. Shoba (2001) asserts that for any dictionary to be regarded as authoritative, linguistic information should be accurately represented. Information on orthography therefore is important in this study as the parallel corpus uncovers the

various spelling variations. The revised orthography surely replaced the old rules; nonetheless, the trails are evident even in recently published documents.

In the preceding sections the concept of lexicography has been discussed widely and it is evident that lexicography is responsible for dictionary making and dictionary research. The overview of isiXhosa lexicography and the features of morphology were discussed.

2.11 Conclusion

This chapter reviewed literature in line with the theoretical frameworks that inform the current research. The first section presented research conducted in South Africa and abroad, specifically on a need for terminological resources that are based on corpus methods. The literature revealed a consensus amongst African scholars, linguists and lexicographers that there is lack of specialised terminology and necessary resources in indigenous languages. The use of corpora proved to be useful both as a tool for terminology extraction and dictionary compilation. In the second section of the chapter, a brief overview of the development of translation studies from equivalence-based theories to descriptive approaches was provided. Before the linguistic era, translation was solely dependent on linguistics and the methods in translating were literal or based on word-to word. The linguistic-based theories came to change those theories. The works of Nida (1964) and Newmark (1981), amongst others brought debates on equivalence and meaning in translation. This period is known as the linguistic era. The central concepts were equivalence and the source text. The target text was regarded as the mirror image of the source text. The notions of dynamic equivalence and communicative equivalence are relevant for the current study as they paved a way towards target oriented translated texts.

In the 1970s equivalence-based theories came under fire because they were too prescriptive and neglected the significance of the target text and its cultural context. Vermeer and Nord's functionalist approach put emphasis on the function and role of the text. Culture became fundamental in translation; this resulted in another shift known as the 'cultural turn'. Functionalism and the cultural turn were geared towards descriptive studies. Holmes' map in 1972 offered a home for translation studies and its internal organisation as a scientific discipline. Of significance to this study are the two

divisions of translation studies into pure and applied linguistics. Dictionaries as translation aids were accommodated.

The further development of Holmes's map by Toury (1995) resulted in DTS as a methodology that improved translation studies. DTS became an inspiration for corpus-based translation studies (Laviosa, 2002).

The foregoing overview of the theoretical framework attests that CTS is informed by corpus linguistics which is an empirical research methodology. The last section outlined the development of lexicography as a discipline interrelated with translation studies. Scholars are in unison that these two disciplines are linked. The link originates from the linguistic principles. The evolution of lexicography from prescriptive methods to a descriptive approach opened a space for corpus-based methods.

The role of corpus linguistics has brought about changes in linguistic investigations. The computers are the root cause of this contribution because without computers electronic corpora would not exist; the speed, consistency, reliability of content are outstanding features. However, the status of African languages and utilisation of corpora, particularly in dictionary compilation, needs to be improved, while the speed at which technology and new terms develop require rigorous innovations. The literature reviewed showed that the information extracted from parallel corpora by means of well-planned corpus tools can improve the inadequate information provided in bilingual dictionaries of English-isiXhosa. An overview of isiXhosa lexicography is the indication that dictionary writing in isiXhosa is not new. It also encompassed term formation and development of orthography rules. The next chapter outlines the research approaches that were used to collect and analyse data.

CHAPTER 3

Research Methodology and Analytical Framework

3.1 Introduction

The previous chapter dealt with theoretical frameworks. This chapter outlines the research methods and processes that were followed in the current study so as to achieve the stated aims. The chapter is organised into six sections: this introduction followed by 3.2 which is the restatement of research questions and aims, while 3.3 describes the research design selected to achieve the aims of the study, 3.4 presents the methods and procedures used to collect data, 3.5 provides a step by step outline of the processes followed in compiling the English-isiXhosa Parallel Corpus and lastly 3.6 offers an analysis of corpus data retrieved from English-isiXhosa Parallel Corpus for the design of a bilingual dictionary of financial terms.

3.2 Research aims and objectives

As indicated in Chapter 1 (section 1.3), the main aim of this study is to explore how parallel corpora can be used in the compilation of specialised bilingual dictionaries of English and isiXhosa. In order to achieve the main aim, the following objectives were identified in order to direct this chapter:

- To create an English-isiXhosa Parallel Corpus of financial texts
- To extract bilingual information from the parallel corpus
- To illustrate how information extracted from parallel corpora can be used to address communicative and cognitive lexicographic functions that will meet the needs of different users of specialised dictionaries.

The study explores the use of electronic corpora as the basis for the compilation of specialised bilingual dictionaries. In order to answer the above questions, a specialised parallel corpus of English and isiXhosa comprising financial texts was designed. Because a corpus cannot be fully exploited without a corpus analysis tool, ParaConc was selected and used to explore the English-isiXhosa Parallel Corpus. It, and its value

for the study, were described earlier. The following section describes and justifies the type of research design selected for this study.

3.3 Research design

Selecting a research methodology and design is a decision-making process that is preceded by a research question and aims: ‘A research design is a plan for a study which provides the overall framework’ (Leedy, 1997:195). In other words, it is a well-organised plan that includes several activities such as methodologies, research sites, instruments and data collection procedures with the aim of answering the research question. The ultimate goal of a design is to provide research outcomes that are valid and credible. Creswell (2014:41) states that ‘research designs are types of inquiry within qualitative, quantitative, and mixed methods approaches that provide specific direction for procedures in research design’. A research that combines a qualitative and quantitative approach is called a mixed methods approach. In addressing the research aims re-stated in section 3.2, a corpus-based approach was selected as a suitable choice for this study because the analysis of the parallel corpus requires a corpus analysis tool. ParaConc, the software selected for this purpose, possesses featues that combine the qualitative and quantintave approaches. Because this study is using a corpus-based method, it is employing a mixed method. Corpus–based translation studies is categorised as quantitative research because it relies on computers and computer analysis software to draw word counts, frequencies and statistics. (Conrad, 1999:3-4) explains further that:

1) Corpus-based studies use a principled collection of naturally occurring texts, i.e. corpus
2) They use computers for analyses
3) Corpus-based studies include both quantitative analyses and functional interpretations of language use.

The statistics on the frequency and quantification of collocations display the strengths of the quantitative approach in the current research. In this study, it is important to state that corpus-based methods do not solely rely on corpus data; other methods could complement further actions. A quantitative approach may include exploratory research

and descriptive statistics, as this study does. Zanettin (2013:31) declares, ‘quantitative and qualitative approaches are radically intertwined in corpus-based translation studies and they are not mutually exclusive’. In this study, the researcher will be using a computer analysis tool in analysing corpus results and applying qualitative analysis where applicable. That is, word counts, frequencies and other forms of statistical information will be presented and interpreted in line with the aims of the study.

The following sections outline the methods and procedures that were followed in collecting data, creation of the EngXhPC and the method used to analyse data.

3.4 Data collection procedures

This section presents the preliminary processes that were followed in the collection of data that was used to create the EngXhPC of financial texts. The preparation of a corpus involves a number of decisions and considerations such as the text types, languages, period covered. The following processes are involved in preparing the design of a specialised corpus.

3.4.1 Identification and location of parallel texts

The first step taken by the researcher was to identify the sources of information that were relevant to the study. The researcher visited the websites of the Eastern Cape and the Western Cape Provinces with a view to identifying the translated financial documents in English and isiXhosa. The two provinces were selected because isiXhosa is predominantly spoken by 78,8% of the population in the Eastern Cape and by 24,7% in the Western Cape (Census 2011). The official languages of the Western Cape Province, for example, are English, isiXhosa and Afrikaans. The annual reports and budget documents published on the respective departmental websites for access by the public were identified, and these were available in the official languages of the Province in terms of the Provincial Language policies. The criteria used in the selection of these texts were:

- The type of financial information contained in the official written documents
- The diversity of texts translated by different translators
- The quality and reliability of translation in the target language, isiXhosa, and
- Accessibility in the electronic format.

The above selection criteria ensured that the texts would not be randomly selected, which would compromise reliability and authenticity of corpus data. After identifying the texts, the researcher was required to follow ethical procedures which are described in the following sections.

3.4.2 Ethical clearance, copyright and permissions

Research ethics focuses on what is morally proper and improper when engaged with participants or when accessing information or data. In consideration of ethical issues two important steps were taken, namely, application for ethical clearance and letters requesting permissions from copyright holders. In compliance with the UNISA policy on research ethics, an application for ethical clearance was submitted to the Research Review committee before undertaking the research. This was done in order to ensure that the 'research is conducted with scholarly integrity, excellence, social responsibility and ethical behaviour' (Unisa Research Policy). The application was approved and ethics approval was issued on 31 August 2015. The ethical clearance certificate is attached hereto as *Appendix 1.*

During data collection, written parallel texts consisting of original texts in English and their translations in isiXhosa were identified in order to create the EngXhPC. After the identification and reading of the texts on the relevant Websites, the researcher was obliged to consider ethical and copyright issues. Before retrieving the parallel texts from the websites and inclusion in the corpus, permission had to be granted by the authors. As noted, this was to ensure that the corpus was 'in good legal health' (Atkins and Rundell 2008:82). All the texts that were uploaded in EngXhPC are in line with research ethics and copyright procedures (cf. section 1.6.2). The copies of both request letters and permission letters are attached at the end of the dissertation as *Appendixes 2 & 3* respectively.

Seeking permission from the publishers was not the straightforward process which had been envisaged. After letters were written to the various departments and institutions, it took time to obtain responses because some of the documents had already been made available on the Web. To seek permission for a document posted for public consumption did not make sense to some copyright holders; hence I had to make telephonic follow-ups to explain why the permission was being requested. On the Treasury website of the Western Cape, only the English source texts were posted; users were requested to write to the department if they wanted either Afrikaans or isiXhosa versions of the original text. This also took time as the Treasury personnel had to retrieve the requested texts and to ensure that they were quality assured before they were sent to the requester.

The data was subsequently collected from Provincial Departments and other institutions, namely, Financial Services Board (FSB) and *Bona* magazine, both located in the Gauteng Province, and Western Cape's Language Committee and Treasury.

3.4.3 Type of documents and selection criteria

The collected texts were written electronic texts in English as the source language and isiXhosa as the target language. The fact that parallel texts were machine-readable made the capturing and text conversion much faster and easier. The types of texts that were selected fall under the following categories: (i) annual reports, (ii) annual financial statements, (iii) information brochures and articles on financial matters, (iv) legislation. The annual reports, budget documents were retrieved as full texts whilst text extracts were taken from terminology booklets or lists. The following is a list of texts that were collected and used to create the EngXhPC:

Annual Reports (AR)

1. Annual Report Western Cape Language Committee 2011/12
2. Provincial Treasury Western Cape
3. Annual Report 2013/14
4. Western Cape Language Committee Annual Report 2013/14

Annual financial statements (AFS)

1. Pan South African Language Board's Financial Statements 2002
2. Provincial Treasury, Western Cape Financial Statements 2011-13

Legislation & Budget Documents

1. Western Cape Adjustments Appropriation Bill 2014
2. Adjustment and Appropriation Bill 2015
3. The Constitution of the Republic of South Africa (chapter 13)

Articles on finance from *Bona* Magazine (2016/17 publications)

1. Master your credit card: April 2017
2. Secret savings: February 2017
3. Invest in yourself: March 2017
4. Clothing account debt trap: January 2017
5. Bonus benefits: December 2016
6. How to prepare for unpaid maternity leave: October 2016
7. Jet-setting on a budget: May 2017

Brochures from Financial Services Board: December 2012

1. Protecting consumers- What is the purpose of the FAIS Act?
2. Collective Investment Scheme Brochure
3. Investment in shares on an exchange
4. The Pension Fund Surplus
5. What is long-term insurance?
6. The Supervision of South African Markets
7. Retirement Funds
8. Why should I have a will? (cf. www.mylifemymoney.co.za)

Government annual reports were presented in five sections/parts covering different topics such as general information, chairperson report, accounting officer's, auditor-general's report and audited annual financial statements. The diversity in the structure pointed to different themes, which has helped in showing how financial language is said and written differently in the same document. The variety of information covered in a single text provides a representative corpus of financial terms written and used in

the official documents. Other documents listed above are publications from the Financial Services Board, a public entity, which is subject to the Public Finance Management Act No.1 of 1999. The brochures selected cover diverse topics in the financial sector, such as: insurance, pensions, wills, capital markets, investment schemes, etcetera. The aim of these brochures is to educate the public.

Bona is described as the only South African publication available in four languages: English, isiZulu, seSotho and isiXhosa (www.bona.co.za). *Bona* publishes issues on how to improve people's lives, from their wardrobes to their finances. The focus of this study is on issues related to money matters. Therefore, the articles selected for inclusion in the parallel corpus inform and equip the readership on how to spend, manage and save money. These articles cover diverse topics and therefore contain terminology used in numerous specific areas of finance management. The common denominator in all these texts is that they are for public interest and can be classified as informative texts. The main function of such texts according to Nord (1997:37) is to inform the reader about objects and phenomena in the real world. The composition of the texts indicates the communicative role they play for the target audience, that of providing information on various aspects of a single field: the domain of finance.

The researcher selected these texts because they are authentic documents that are prepared by recognised institutions and posted on the official websites which are reliable. The translations are always performed by professional translators and quality controls are conducted before publication. As Bowker and Pearson (2002:54) assert, 'texts that are produced by authors with proven credentials are likely to contain more authentic examples of LSP use than texts by authors who are not proven experts'. In other words, as mentioned, authorship has an influence on the reliability and authenticity of the whole study. The element of authenticity is further strengthened by the fact that the selected texts were not developed for the purposes of this research but are documents that were published for genuine communicative reasons.

3.4.4 Corpus size

One of the important issues to be taken into consideration when designing a corpus is the size. The issue of corpus size is controversial because in the corpus design literature there is no consensus on the minimum and maximum size. However, the size of a corpus is guided by the corpus typology and the aim of its design. The size of a specialised corpus cannot be equal to the size of a general-purpose corpus. A special purpose corpus generally is smaller because it is specific and deals with a special subject field, which is restricted. The EngXhPC under investigation, covers a restricted specialised language written and used in financial documents published by government and public institutions. Therefore, it is not the intention of the researcher to compile a large corpus due to scarcity of translated texts in the chosen subject-field. Many scholars attest that specialised corpora are usually smaller in scale than general language corpora because of their narrow focus.

The EngXhPC design was informed by two reasons: the availability of translated financial texts and the ultimate goal of the study. The translation of financial texts from English into indigenous South African languages is still a new language practice that is posing challenges. Bowker and Pearson (2002:45) caution that 'you may find that you can get more useful information from a corpus that is small but well-designed than from one that is larger but is not customised to meet your needs'. The researcher made a careful selection of texts that contained the kind of terminology which might be useful in compiling a representative parallel corpus that suits the purposes of this study. The EngXhPC can be described as a small specialised parallel corpus with a total of **154 029** items (English 88 101 and isiXhosa 65 928) as illustrated in Figure 3.5 below.

3.4.5 Text types, period coverage and authorship

The types of texts that were selected for inclusion in a corpus are always determined by the type of corpus to be designed. The EngXhPC as a specialised parallel corpus was designed for purposes of extracting technical terminology for dictionary making purposes. Technical terms can be found in different genres such as legal, medical, science and technology, etcetera. The parallel corpus in this study focuses on the domain of finance. In other words, it is a single-field corpus restricted to one domain rather than a wide range of technical or specialised domains. The texts that were

included in the EngXhPC cover a period between 2002 up to 2017 and were translated by a variety of translators contracted by different departments or organisations. The recent publications were selected in order to capture the latest terminology in the field under investigation. Language for specific purposes is very dynamic; therefore capturing old terminology would not be beneficial to users of specialised dictionaries.

The table below displays the characteristic features of the EngXhPC compiled and its suitability in being a source of lexicographic data required of a specialised bilingual dictionary of finance.

Table 3.1: Illustrating the features of the English-isiXhosa Corpus

Corpus type	Specialised parallel corpus
Domain (single-field)	Finance
Directionality	Unidirectional
Function	Encoding
Number of languages	English-isiXhosa
Quality	Text translated by professionals
Time	Contemporary

The EngXhPC can be described as a unidirectional parallel corpus of financial terms, which is contemporary. A well-constructed corpus is a recipe for successful analysis.

3.4.6 Challenges encountered during data collection

This section outlines the challenges that were encountered during the collection and capturing of corpus material. As discussed in section 3.4.2, requesting permission to use data for research purposes led to its own difficulties. Obtaining permission from financial institutions was problematic due to their red tape; as such, data from the private sector was not included in this corpus. One of the constraints during the selection of parallel texts was choosing ‘financial information or annual financial statements’ only to discover they had not been translated, because according to the

publishers translating financial information is cumbersome and costly. Documents that had no translations were not captured since this is a parallel corpus.

Initially, the provinces that were selected to collect parallel texts were the Eastern Cape and the Western Cape because isiXhosa is one of the official languages spoken by the majority of citizens. However, when searching various websites, no Xhosa versions of the annual report or budget documents were posted on the Eastern Cape Provincial official websites, and this was confirmed telephonically. This limitation does not affect the corpus balance and representativeness. The three categories of texts retrieved for the creation of a parallel corpus focusing on financial terms provide adequate information to achieve the aims of the current study. The translations of these documents by different professional translators in various locations enhance the reliability of financial terminology. The subsequent section outlines the type of corpus data extracted from EngXhPC.

3.5 Compiling and analysing the English-ISIXhosa Parallel Corpus using ParaConc

The analysis of a parallel corpus is preceded by various steps ranging from text selection to corpus alignment which is a precondition of a parallel corpus. This section, therefore, describes the steps followed in the compilation of the said parallel corpus.

3.5.1 Pre-processing of corpus data

The parallel texts that were collected were in different formats, which necessitated text conversion and clean-up. Most texts were posted on the website in a pdf format, a format that is not editable. All the texts were then converted from pdf to MSWord documents in order to allow text normalisation. Text normalisation involves basic processes such as cleaning of data, correction of spelling mistakes, and removal of characters that emerged during conversion of texts from pdf to Word. The text conversion was automated by a Microsoft office pdf converter while the clean-up was done manually by the researcher. This process was the initial cleaning of documents in MSword. It should be noted that there were no changes made to the original content of the texts except the removal of tables, pictures and graphs, as these are not readable in the parallel corpus which is word based.

After the completion of these basic processes, the texts were then converted from Word to plain text (.txt). Besides the fact that, technologically, plain text uses a lesser storage space than rich text formats, all corpora according to Bowker (2003) are text-based resources, just like the EngXhPC. The plain text is also editable and the researcher was able to perform quality checks before the files were loaded onto the corpus. Cleaning of plain text files was done manually in order to eliminate 'noise' in the corpus. 'Noise' in a corpus simply means unnecessary information such as misprints, or strange characters that might affect corpus analysis. It should be noted that ParaConc does not permit cleaning up or correction of any mistakes, so all the errors should be removed from the plain text. Figure 3.1 and 3.2 below illustrates bilingual plain texts in the source and target language.

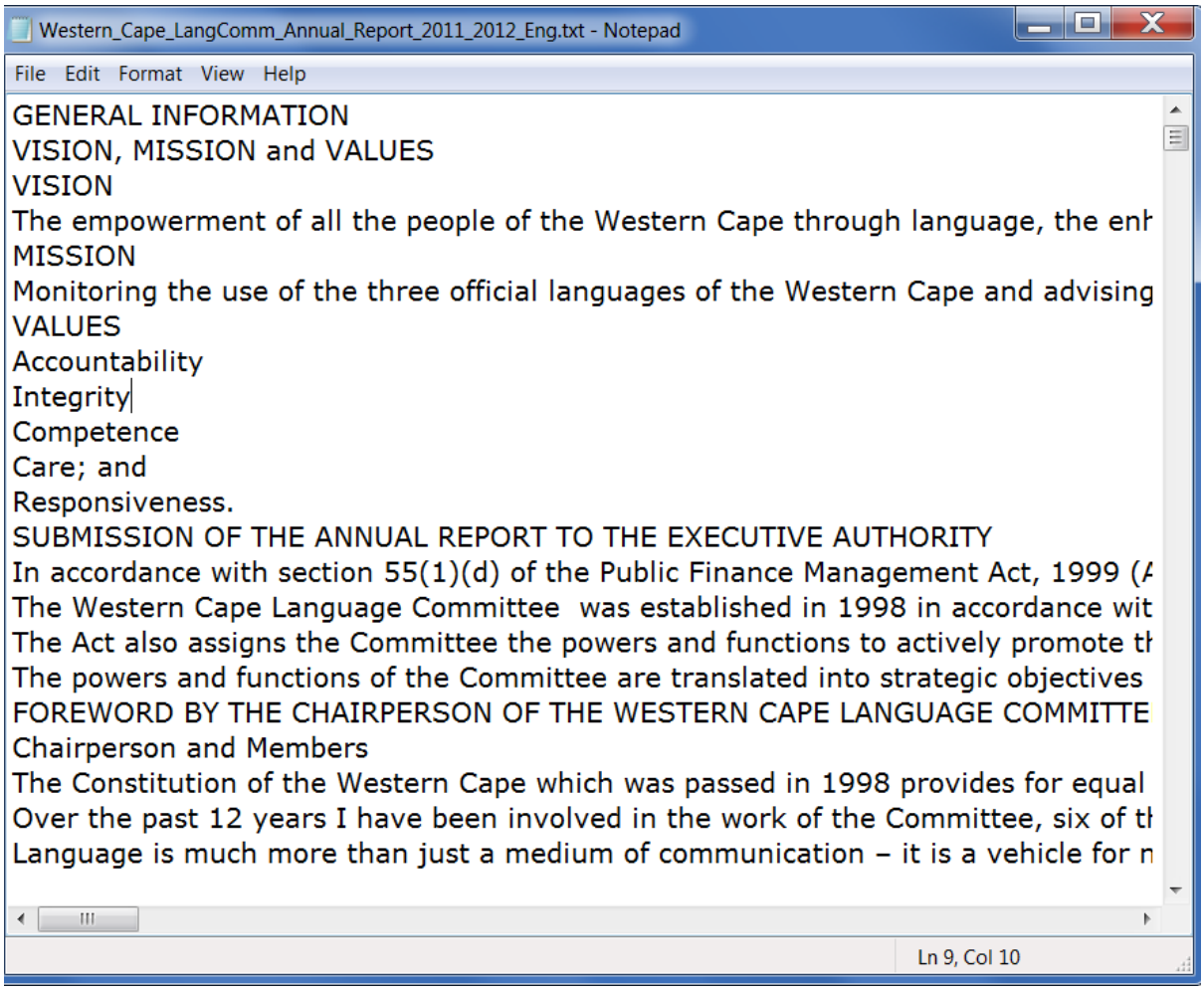


Figure 3.1: A plain text format in English

Figure 3.1 is an illustration of a plain text format in the source language. It was mentioned above that txt. format occupies a lesser space than a rich text. The format runs horizontally from left to the far right of the screen. Figure 3.2 below illustrates how the plain text of the target text, isiXhosa, looks after normalisation.

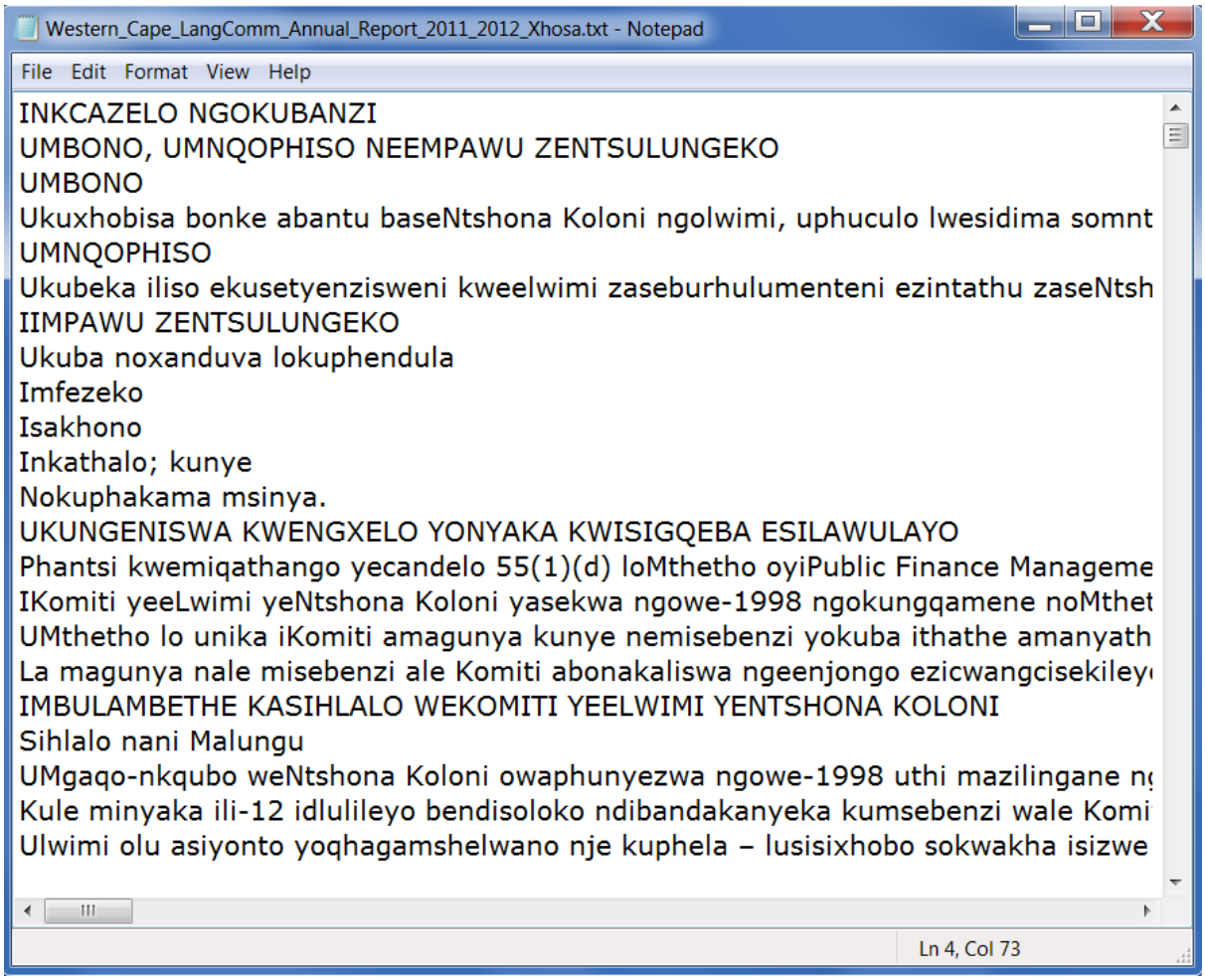


Figure 3.2: A plain text format in isiXhosa

Figures 3.1 and 3.2 are the examples of files that have been converted to a plain text and are ready to be loaded on ParaConc. After the files in the plain text format were cleaned up by the researcher, they were ready to be uploaded on computer software.

3.5.2 Loading of corpus files

To upload the .txt corpus files, the software program was opened by clicking on ParaConc.

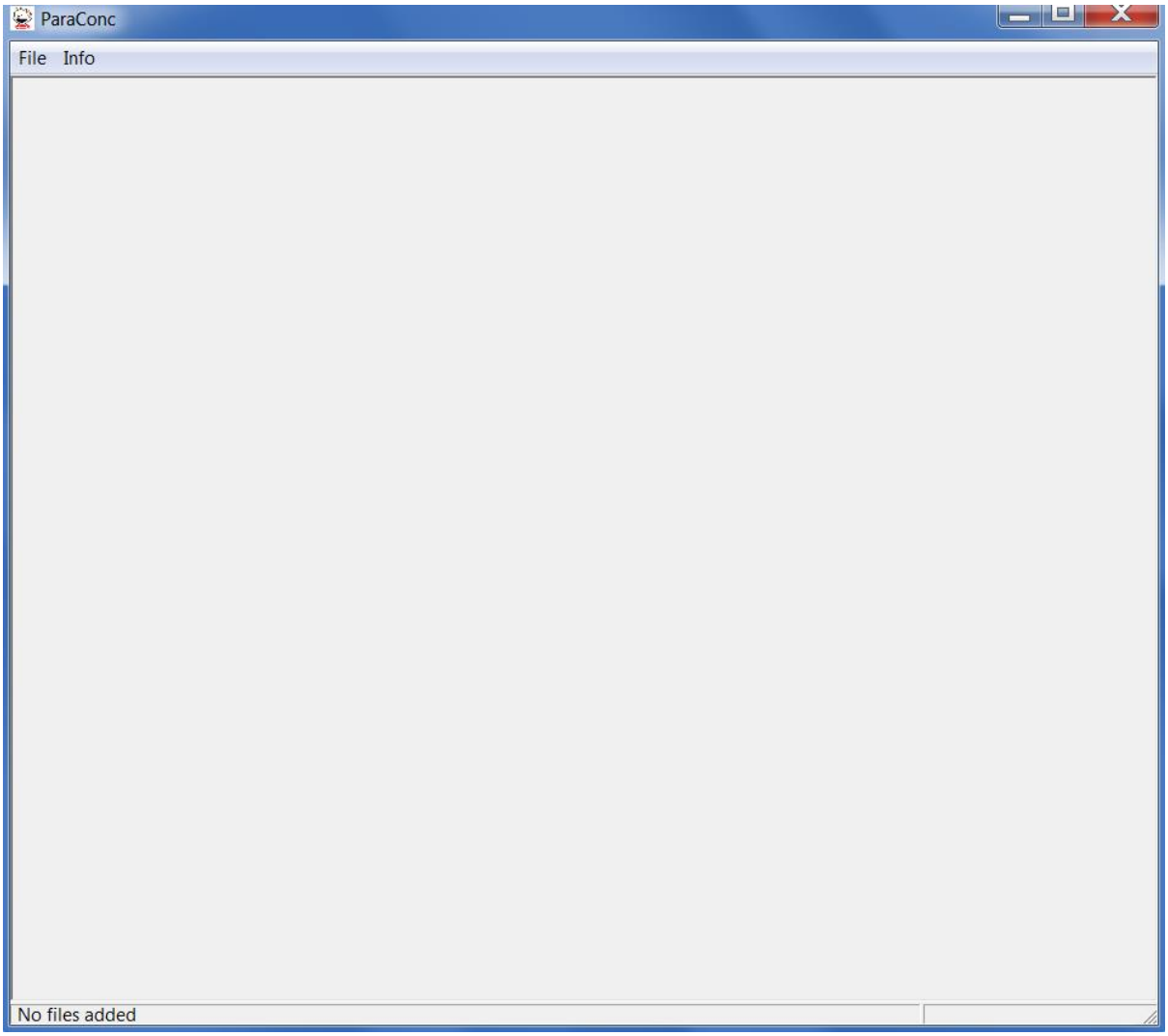


Figure 3.3: Initial screen

After opening ParaConc, an initial screen with two menu items: FILE and INFO appeared (see Figure 3.3). The diagram shows a blank window. This is confirmed by the information field in the left lower corner of the window which reads:

'No files loaded'.

Before loading the files, a file menu bar on the left window was double-clicked to open ParaConc; then a dialogue box appeared, enabling the selected parallel texts to be loaded. A load corpus command was selected amongst others as shown in Figure 3.4.

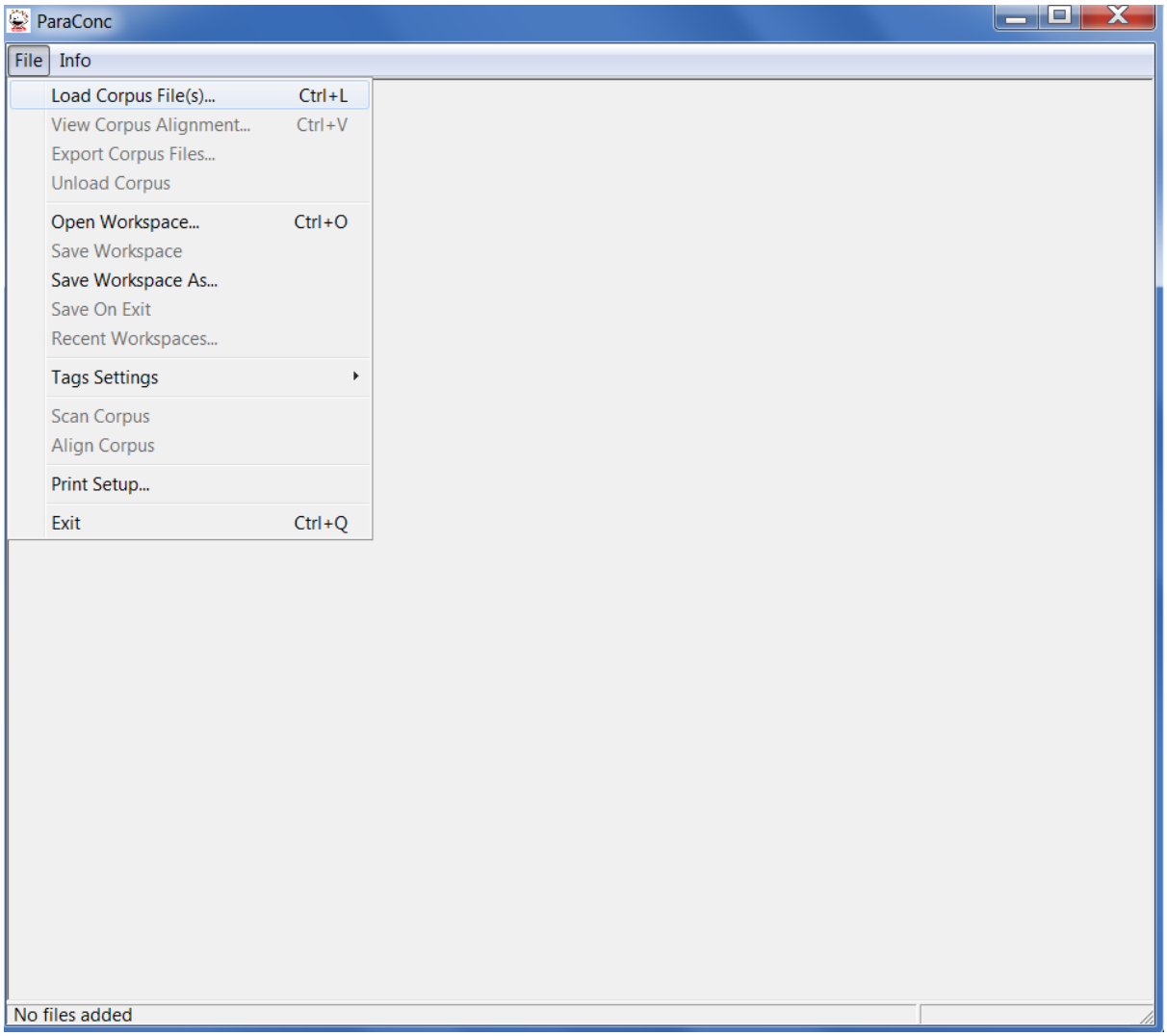


Figure 3.4 Loading corpus files

After a load corpus file command was shown, a dialogue box appeared to enable the selection of parallel texts. Before adding the files, the source language and the target language were selected.

The source language in the corpus under compilation is English; therefore, English was chosen while the target text was isiXhosa. When running the down drop box, isiXhosa was not listed; fortunately, ParaConc allows one to choose any language or font that can represent unlisted target languages. In this instance, French (Canadian) was selected to represent isiXhosa as shown in Figure 3.5 below.

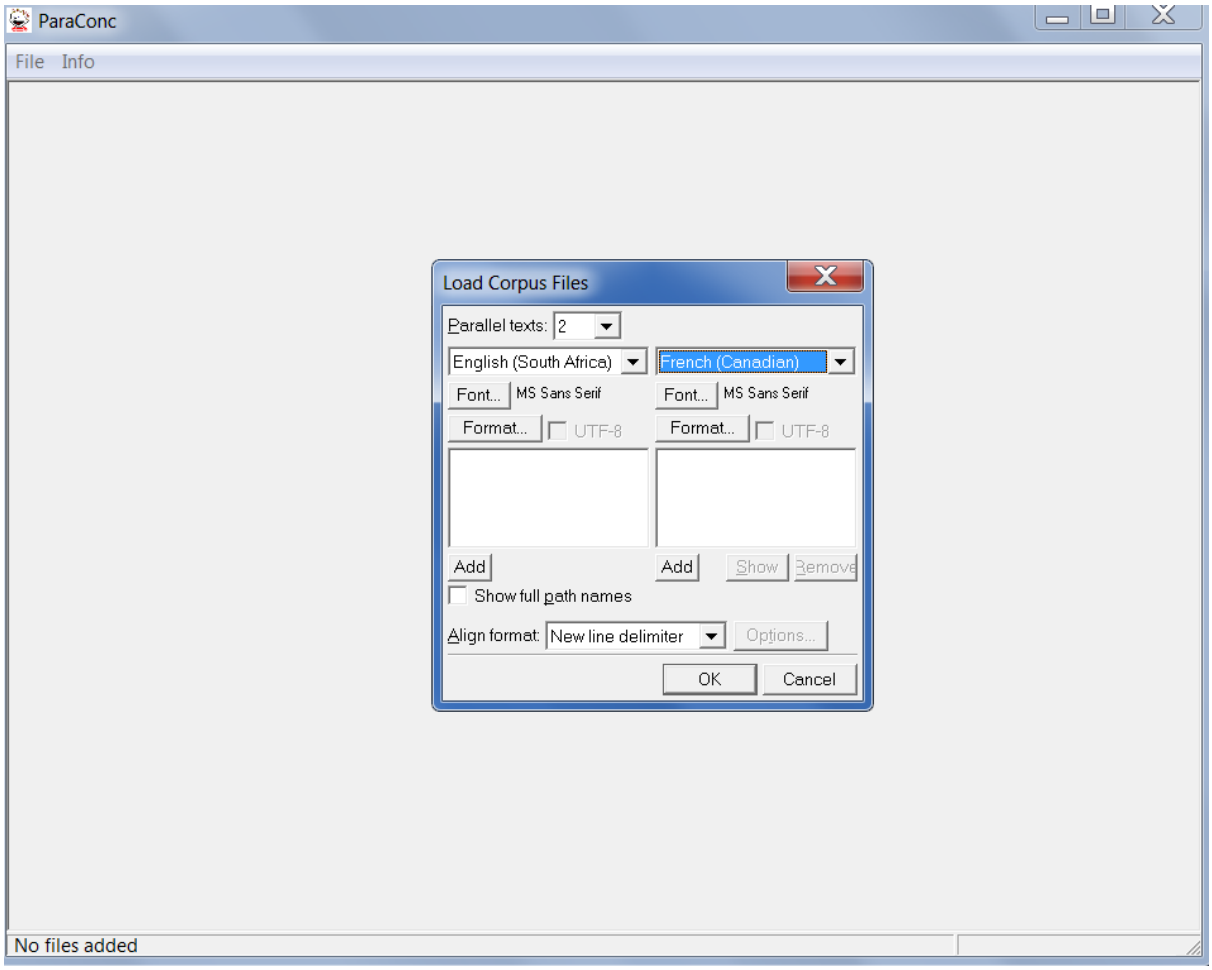


Figure 3.5: Selecting files and languages

All the files were loaded by clicking on the Add button. To finish loading corpus files, the compiler clicked on OK at the bottom of the dialogue box and the two languages were captured as shown in the diagram below.

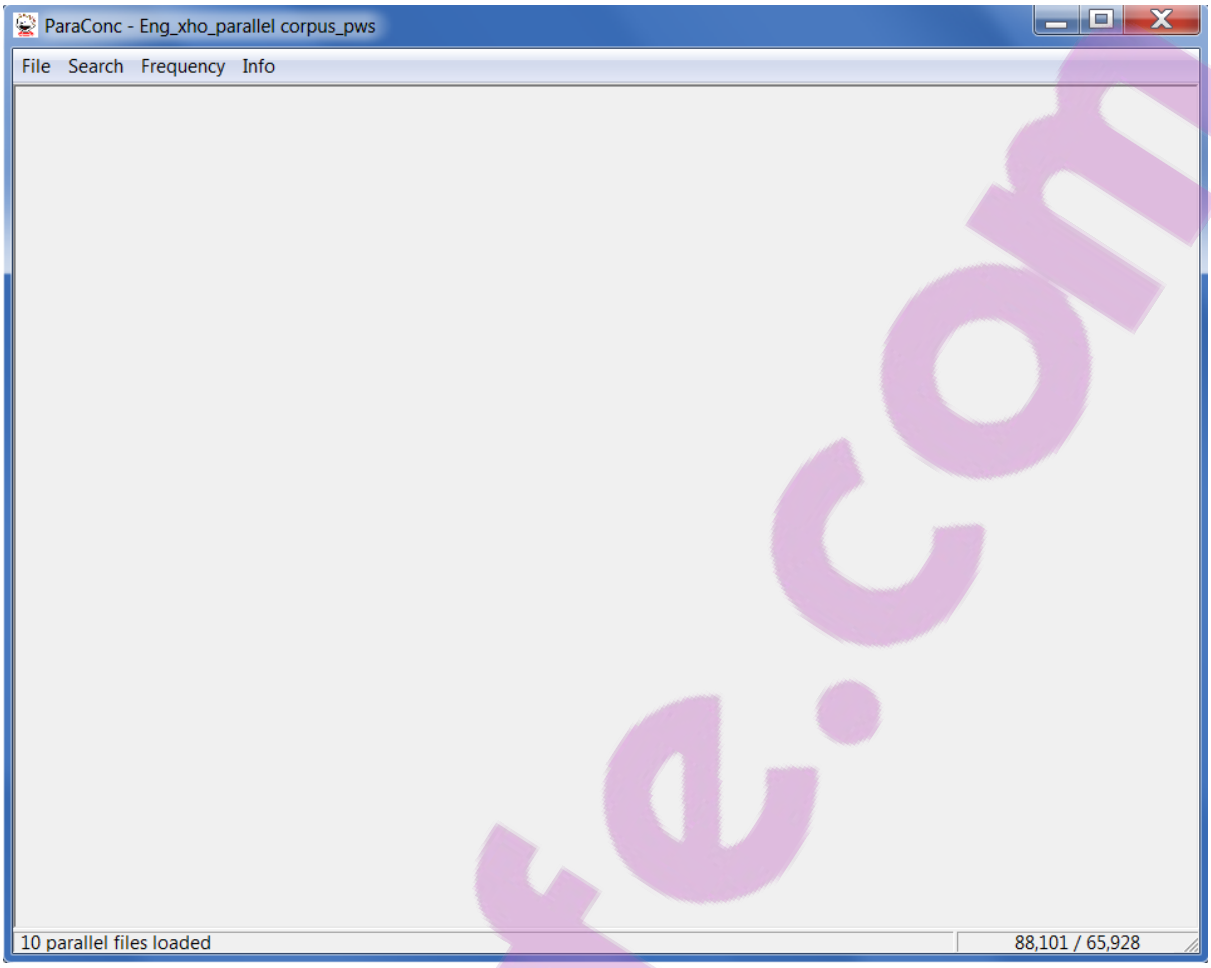


Figure 3.6: A view of EngXhPC files uploaded into ParaConc

After loading was completed, the number of files loaded appeared on the left of the menu, whilst word counts were displayed on the far right of the menu window (see Figure 3.6 above). The following diagram shows the number of the loaded parallel files.

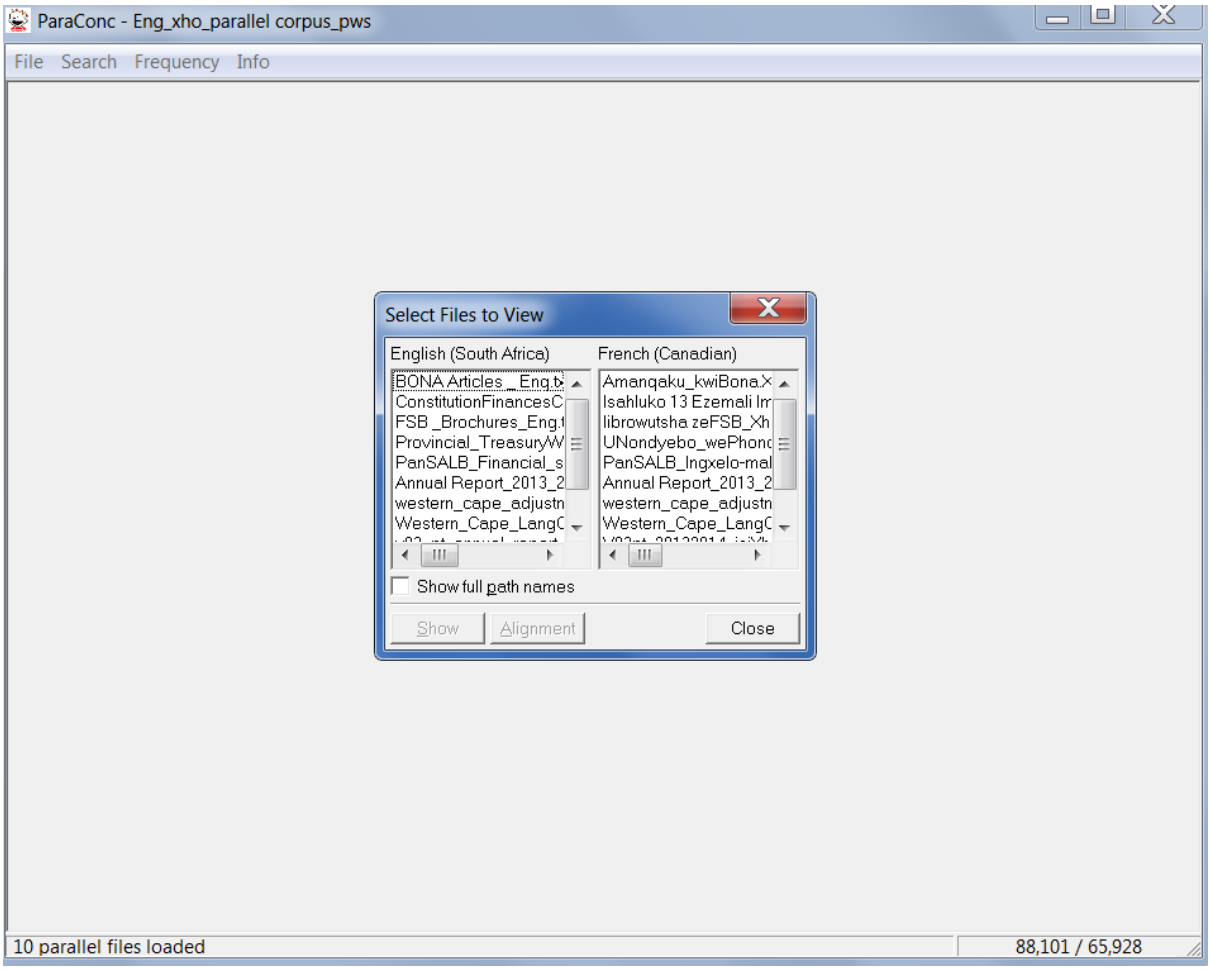


Figure 3.7: Parallel texts loaded in the workspace

Once a corpus is loaded, the menu items like **file, search, frequency, and info** appear on the menu bar. These items facilitate the exploitation and analysis of the parallel corpus. The window in the centre displays the file names in their alphabetic order. Figure 3.7 above illustrates the number of corpus files loaded. The number of files loaded in the English-isiXhosa Parallel Corpus are shown on the left-hand side corner with a total of 88 101 English words and 65 928 isiXhosa shown in the right-hand side of the screen shot. There is a notable difference between the two sub-corpora, because English is written disjunctively whilst isiXhosa is written conjunctively. This type of disjuncture will be noticed throughout the study. An explanation of the alignment process and its significance thereof now follows.

3.5.3 Alignment

In order to proceed with corpus querying, corpus files had to be aligned. The process of creating links between the texts is generally referred to as alignment (Bowker and Pearson (2002). The sentences in the source text are matched with the sentences in the target text. This is a crucial step when compiling a parallel corpus. The correspondence between sentences enables ParaConc to function effectively. In aligning the loaded texts of English and isiXhosa, the compiler clicked on the **File** menu and then moved to another icon: View the Corpus Files. The load corpus files box appears (see Figure 3.8) with loaded files.

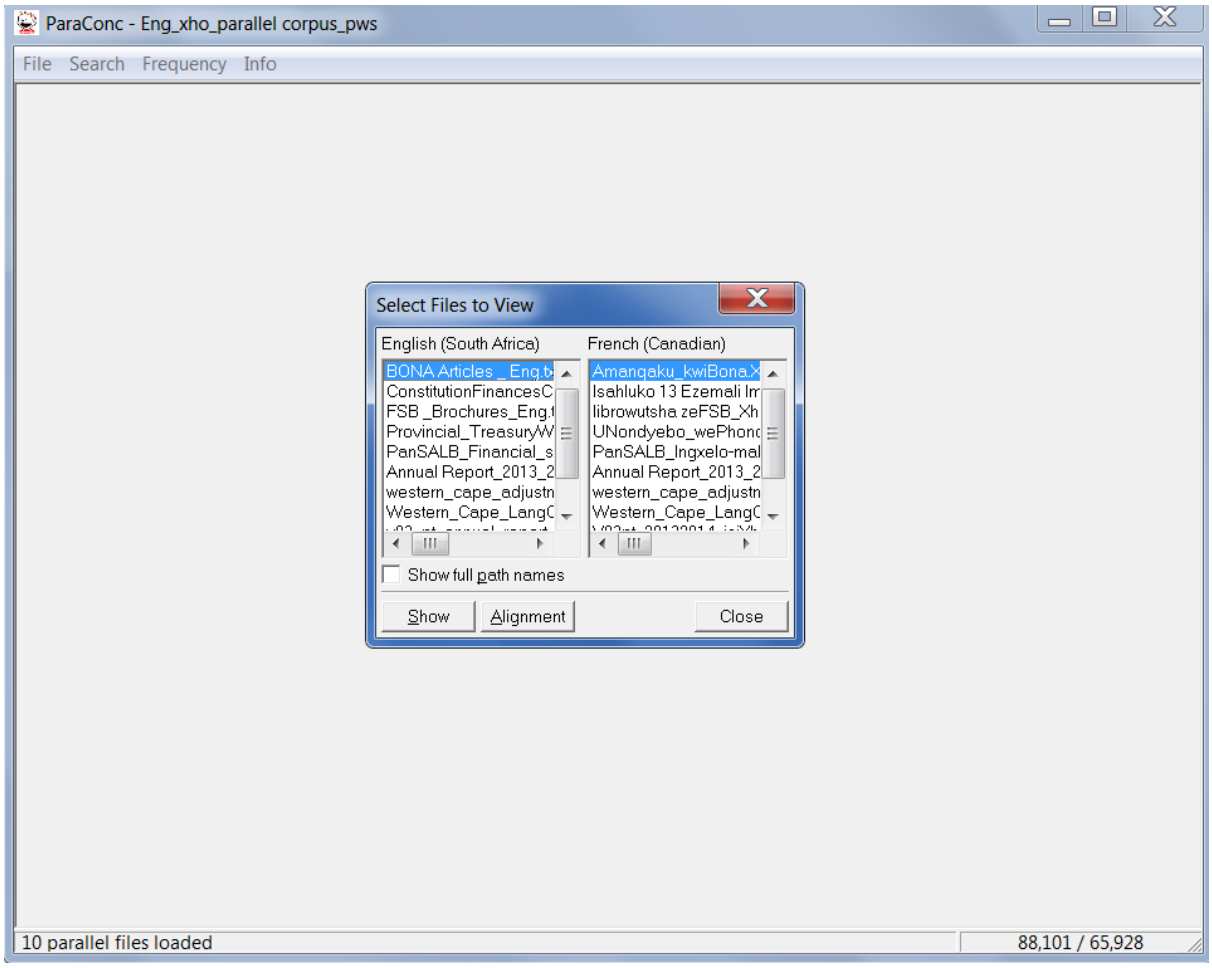


Figure 3.8: English and isiXhosa files selected for alignment

After right clicking on the parallel texts, both the source and the target texts are highlighted indicating that they are active. A click on Alignment enabled ParaConc to automatically run the alignment.

The alignment utility by ParaConc was not fully automated, therefore, human intervention was required. Figure 3.9 below displays the semi-automated alignment by ParaConc.

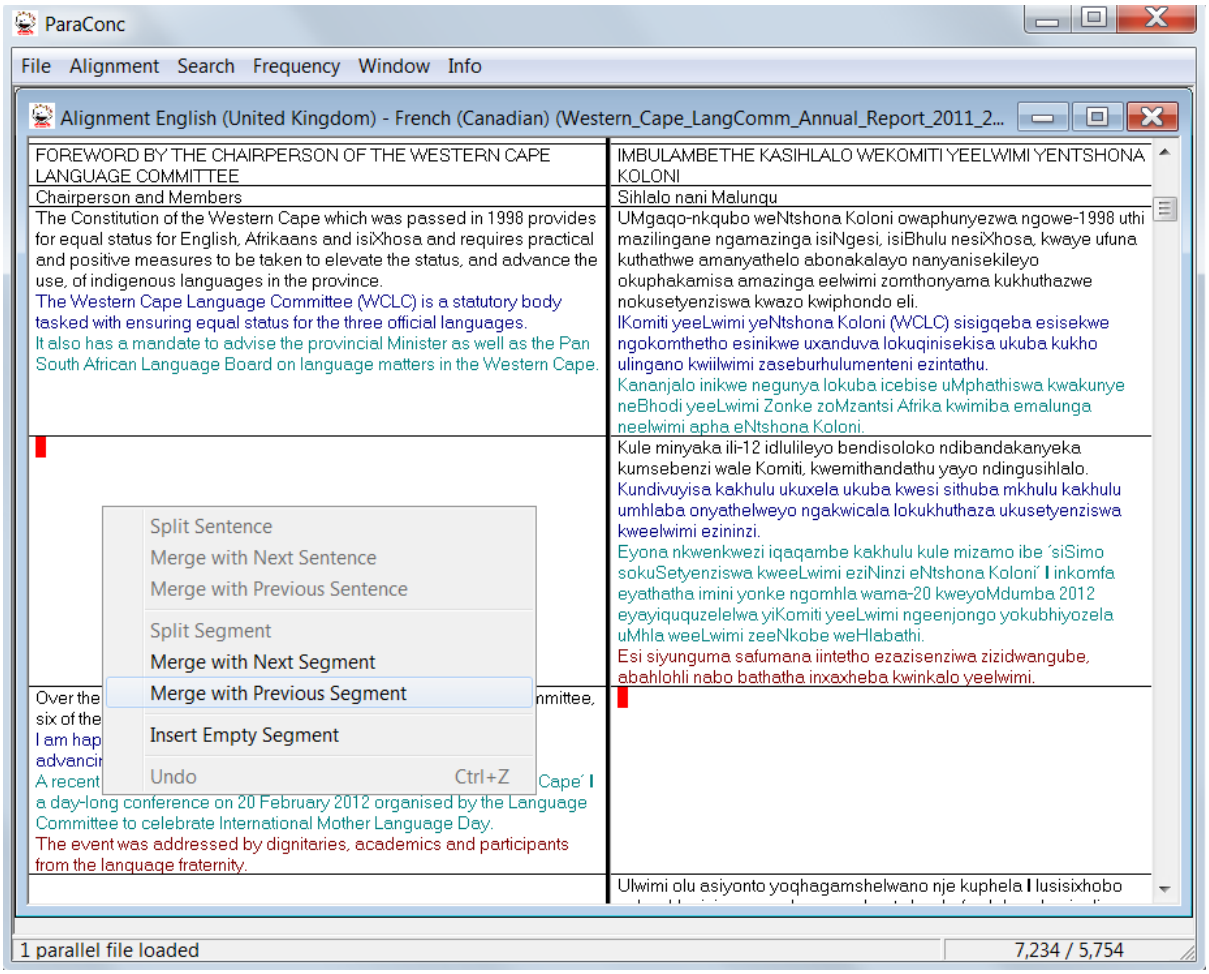


Figure 3.9: Semi-automatic alignment of English-isiXhosa texts by ParaConc

ParaConc has a window-based aligner that performs semi-automatic alignment as shown in Figure 3.9 above. Corpus alignment can be performed at different levels, word, sentence and paragraph. The parallel texts in Figure 3.9 above are in sentence segments. The blank spaces in the window indicate that some segments were not aligned by the software and this necessitated a manual alignment by operating other functions namely, **merge with next segment**, **merge and split**, **insert empty segment.** This was one of the time-consuming exercises that needed analytic skills and linguistic competence. In order to link segments, the researcher had to select an appropriate option from the alignment box. In the example above, an empty segment had to be linked with the one on top, and **merge** with previous segment function had to be performed by a right click on the first word of the segment to be moved. 'The alignment is the indication of equivalent text segments in the two languages' (http://www.athel.com/para.html (accessed on 11/05/2016)). A sentence is the basic alignment segment in ParaConc (see Figure 3.10). The colour coding is a functional feature that links the source text segments with the target sentences. If an original segment is green its counterpart displays the same colour. Text alignment plays an important role in facilitating the retrieval of bilingual concordances. After a satisfactory alignment as displayed in Figure 3.10 below, the aligned texts were saved as workspace.

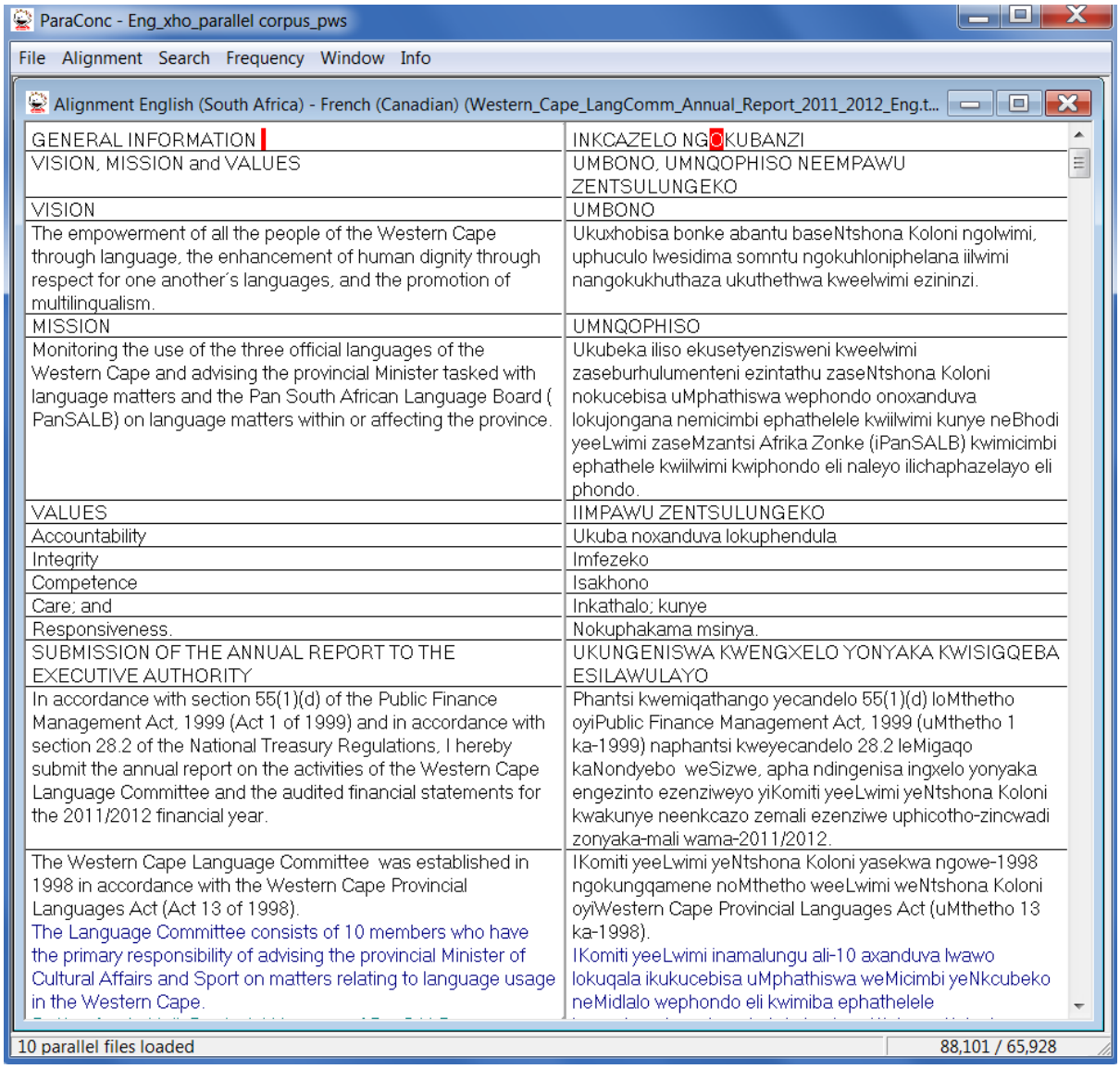


Figure 3.10: Aligned parallel segments: English and isiXhosa

Saving aligned texts on the workspace keeps them intact and allows for further interrogation without reloading the files. The EngXhPC was saved as a special ParaConc Workspace file (.pws).

A folder to hold all workspace files was opened separately from other corpus files. Once the corpus was saved, a search menu appeared with items: File, Frequency, and Info with all different options of search. The next sections will explain and demonstrate how ParaConc was used to perform various functions that are important in dictionary-making in the EngXhPC.

3.5.4 Alphabetic word list

ParaConc can generate word lists in alphabetic order in both the source language and the target language. This feature is important in dictionary making because it makes it easier to identify search words. In the **frequency menu,** the information is selected by clicking on the option **Alphabetic order**. The various options open, to generate an alphabetic list on either one language or both. The alphabetic word lists in both languages are of great interest in the current study; therefore two lists were generated by selecting the command **ALL** English and then French (Canadian) for isiXhosa in the drop-down box.

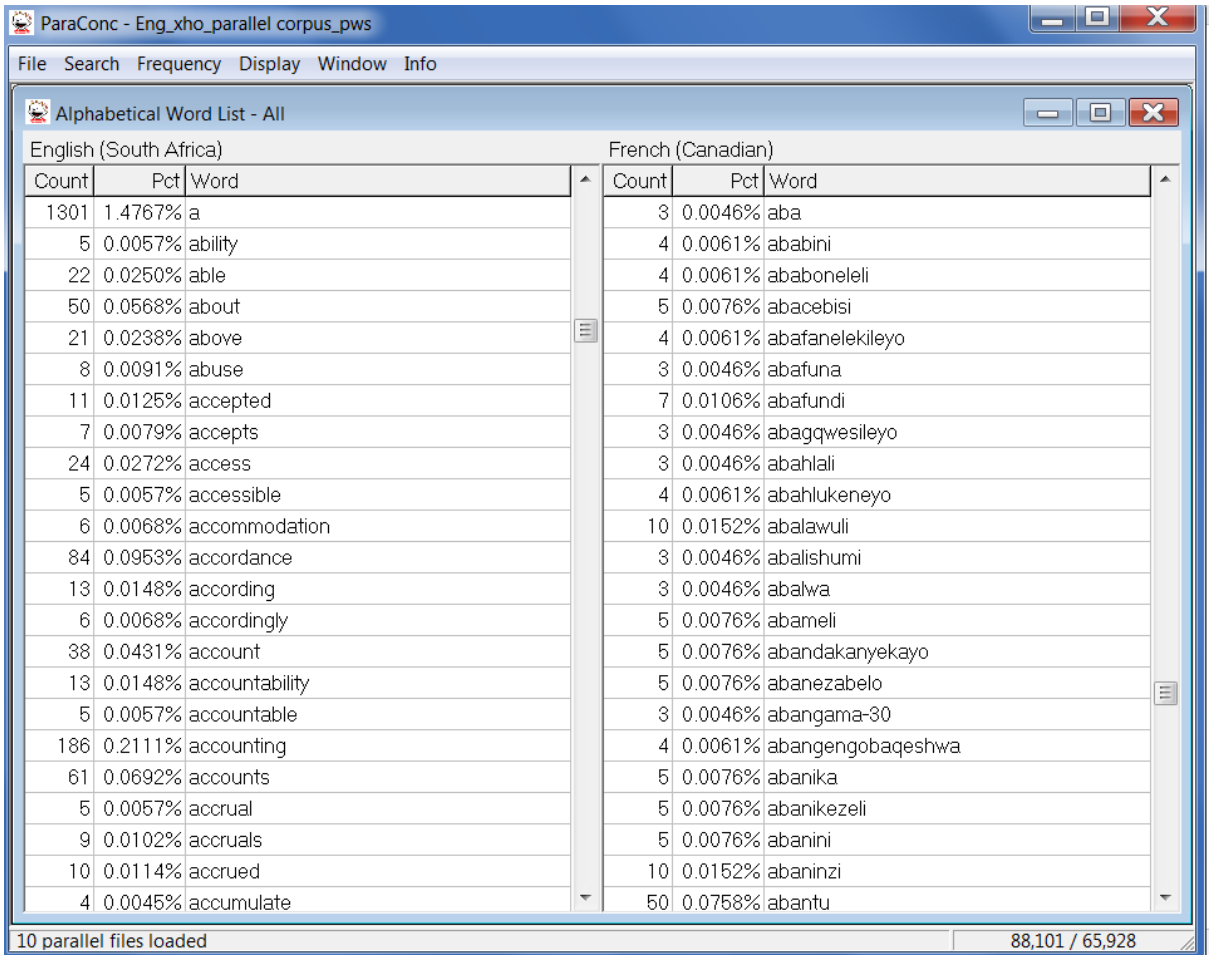


Figure 3.11: English-isiXhosa alphabetic word lists

The two alphabetic word lists generated by ParaConc above provide information on each word away from its context and give an indication of the words that are available in the parallel corpus. Both lists indicate all the words that begin with letter ‘a’ in the parallel corpus. The English alphabetic list begins with ‘a’ whilst isiXhosa starts with a

demonstrative *aba.* This shows that the order reflects alphabetic arrangement and does not mean that the words are corresponding in the two lists. The statistical information in the form of word counts is also indicated before each word. The alphabetic lists in dictionary making assist in building a headword list according to alphabet, identifying possible headwords, their structure and the spelling as shall be shown in the next chapter.

3.5.5 Creating a frequency list

One of the major advantages of using corpora in lexicography is to obtain information regarding word frequency. One of the basic features of ParaConc is its ability to provide frequency lists**.** Bennet (2010:4) defines a frequency list as a list that 'displays the words occurring in a corpus along with the number of times each word appears'. Frequency lists focus on the most frequent words in the corpus and arrange them accordingly. After the alignment process that is explained above, the first search was aimed at discovering the number of words that exist in the EngXhPC and how many times each word appears. To create a frequency list, the researcher clicked on **Frequency** on the menu bar and then selected frequency order. Frequencies can be investigated from the whole corpus or on individually aligned texts. An analysis from the whole corpus was the best option because the researcher wanted to know which words are frequent in the EngXhPC. To retrieve this information, the command: **All** was executed to create English and isiXhosa frequency lists (see Figure 3.12 below).

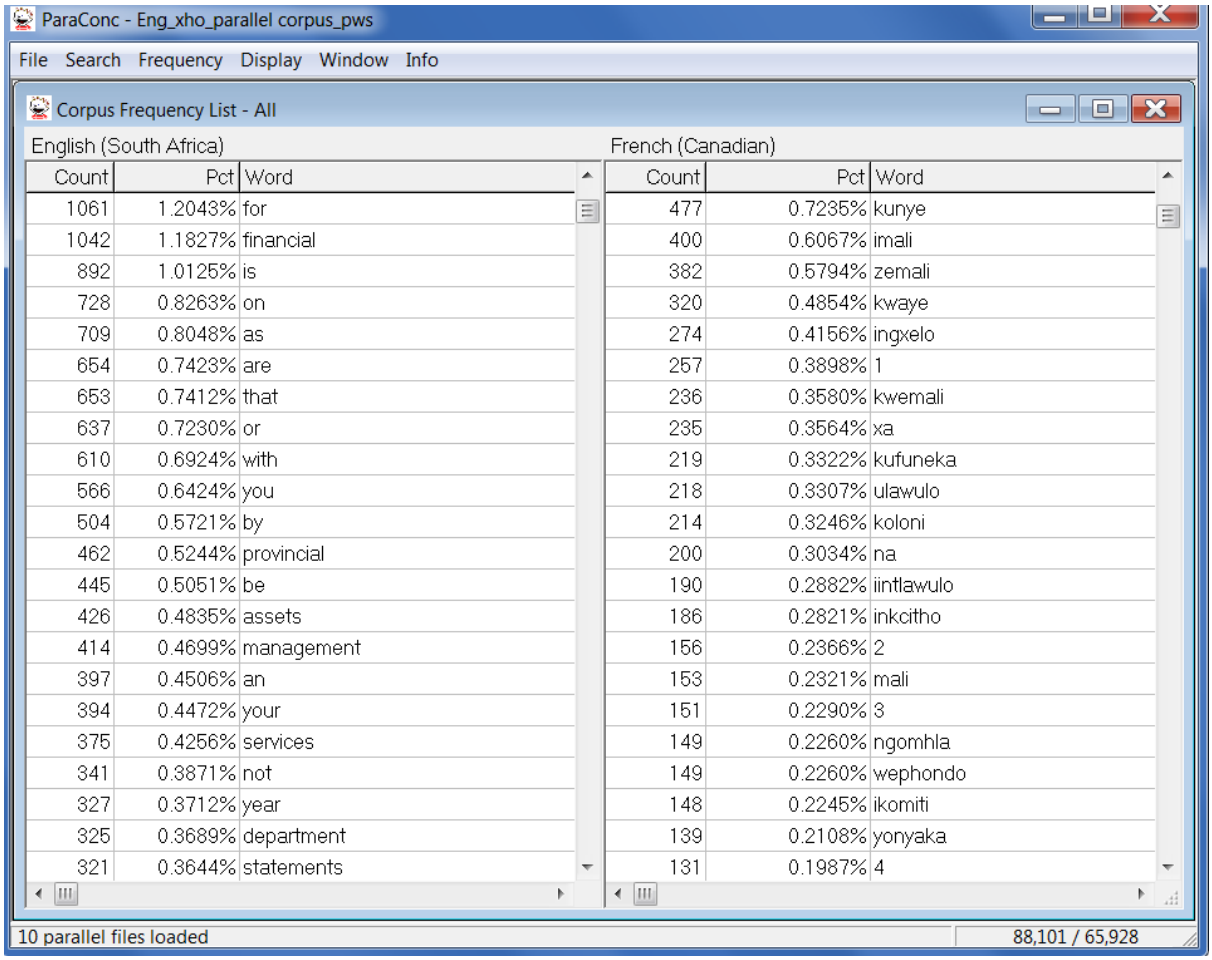


Figure 3.12: Frequency from English-isiXhosa Parallel Corpus

The frequency order search reveals that the preposition ‘for’ is the most frequent word in the English corpus with 1061 counts, whilst *kunye* is highly frequent in the isiXhosa corpus. The two words are function words which are not usually lemmatised in specialised dictionaries. However, going down the frequency list of financial terminology, the term ‘financial’ has 1042 counts in the English order while *imali* in isiXhosa ranked high (with 400 frequency counts as highlighted in blue. ParaConc has the ability to remove all the words that are not relevant in the frequency list. As observed above, a frequency list is arranged according to the number of occurrences in a parallel corpus. The problem of deciding what to include and exclude has been a major challenge in lexicography, already discussed in prior studies (Tymoczko, 1998; De Schryver and Prinsloo 2002; Ndhlovu, 2016). Frequency lists aid in the selection of headword lists. The frequency lists in Figure 3.11 above show common words used in the parallel texts. The higher the frequency count of a word, the higher the chances of its inclusion in the dictionary. The inclusion of frequent words also ensures that words

that are no longer used in the language are not included, thus giving space to more frequently required words such as new ones. A parallel corpus can be investigated by performing text searches which constitute a crucial feature in parallel investigations. The next section explains how ParaConc can be exploited in doing various searches in the EngXhPC.

3.5.6 Search Function

To initiate a search, the researcher selected **SEARCH** from the search menu. After clicking on a search option, a search menu box appeared allowing for possible searches (see Figure 3.13).

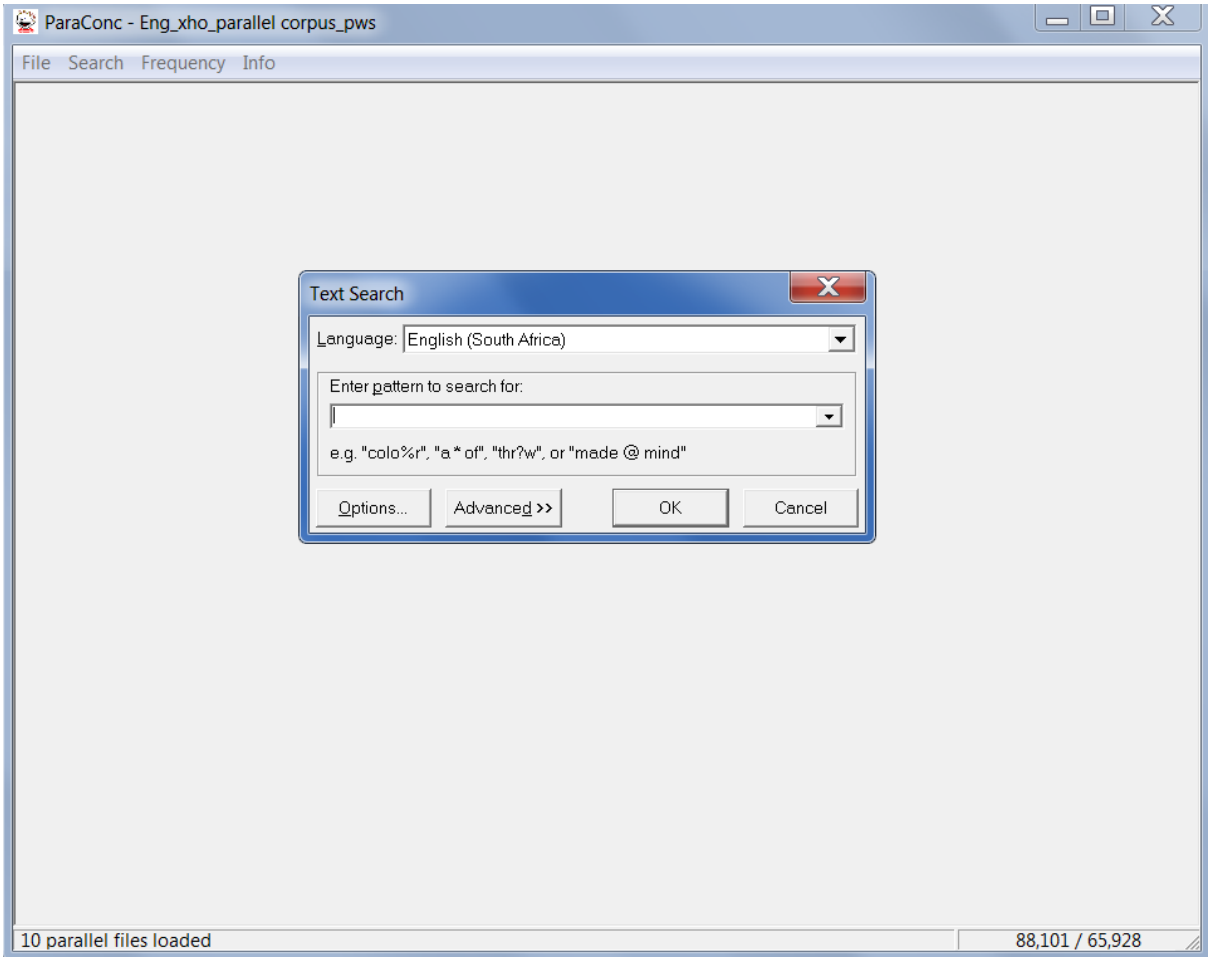


Figure 3.13: Text search menu

In preparation for a search in English, the language to be queried was selected from the language dialogue box as demonstrated above.

To perform a simple search, the researcher then typed the search term ‘financial’ in the search pattern box as displayed in Figure 3.13. The selection of the search patterns was guided by the frequency order in the frequency lists. In the frequency lists in Figure 3.12, ‘financial’ was the second highest ranked word with 1042 word counts.

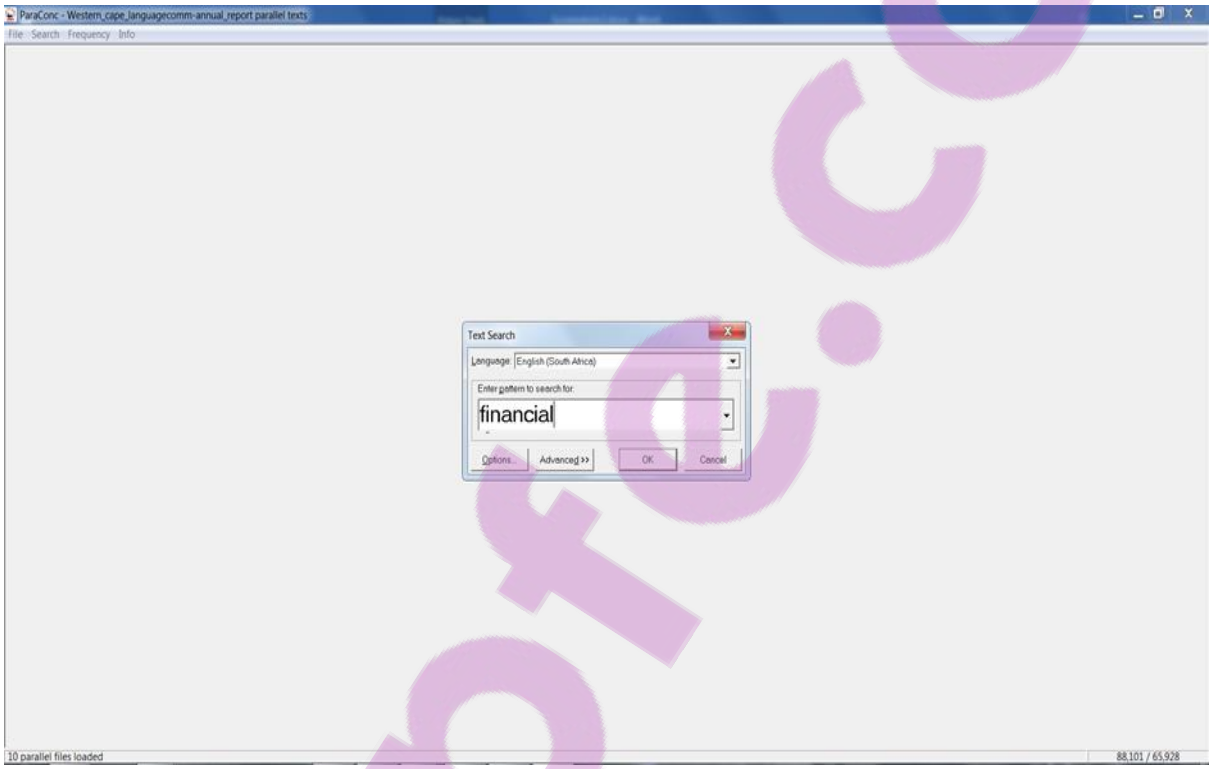


Figure 3.14: Performing a simple search of ‘financial’

After a click on OK, the software performs the search. The results of the search in Figure 3.14 show that the two languages are separated by a dividing line; in the former English appears in the upper window and the translations in isiXhosa are in the lower window below the dividing line, whilst in the latter it is vice versa. The details appearing down the concordance windows provide us with the language of investigation, the string matches and the number of matches in the whole corpus. The results of the text search are always displayed in the concordances which are discussed below.

3.5.7 Displaying concordances

When a single word or phrase is searched, a concordance is displayed where concordance information is presented in two windows. The top window is representing the results of the search whilst the second displays the possible translations of the sought word. A concordance can be defined as 'a display showing all the occurrences of your search patterns in their immediate contexts' (Bowker and Pearson 2002:230). ParaConc contains a fascinating feature that generates certain strings or patterns in a parallel text or corpus. To generate the concordance information from the loaded corpus, a search menu is opened and a click on the search option displays a text search box which allows one to select the language on the upper window. To retrieve the concordance lines of the word 'financial', the term was entered in the search box and clicked to allow the program to search the word as shown in the concordance window below. This resulted in the concordance lines with the search word highlighted at the centre of the upper window whilst the lower window displayed the concordance lines in isiXhosa, the target language in the corpus as shown in Figure 3.15 below.

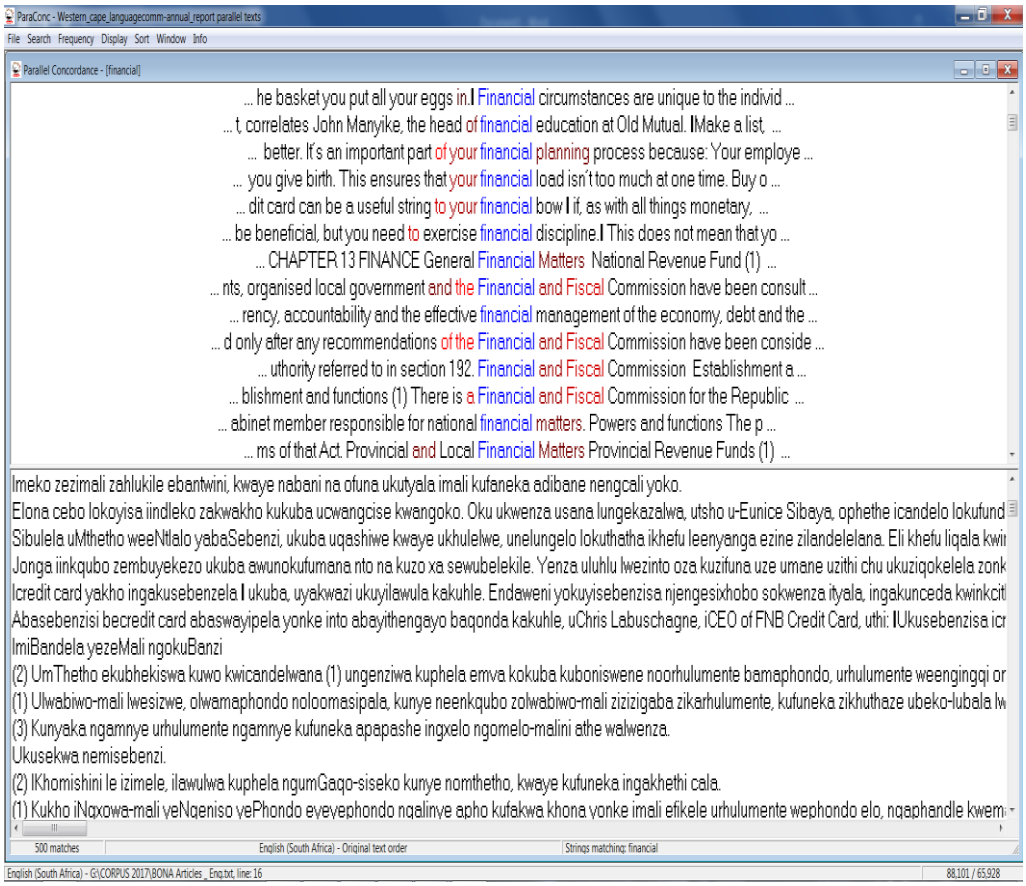


Figure 3.15: Concordance information of the term 'financial/-*mali*'

In the diagram above, the search term ‘financial’ is displayed in one colour in the centre of the concordance window. It is further noticed that the node is surrounded by other words that give context to the individual term, hence the format is called Key Word in Context (KWIC). Words that co-occur with the centred word were also highlighted to indicate potential collocates.

The software also allows a search in any languages in the corpus. To search for a term in isiXhosa, under language box, French (Canadian) was selected and the search term *imali* was entered in the text search box to yield the results as shown in Figure 3.16 below.

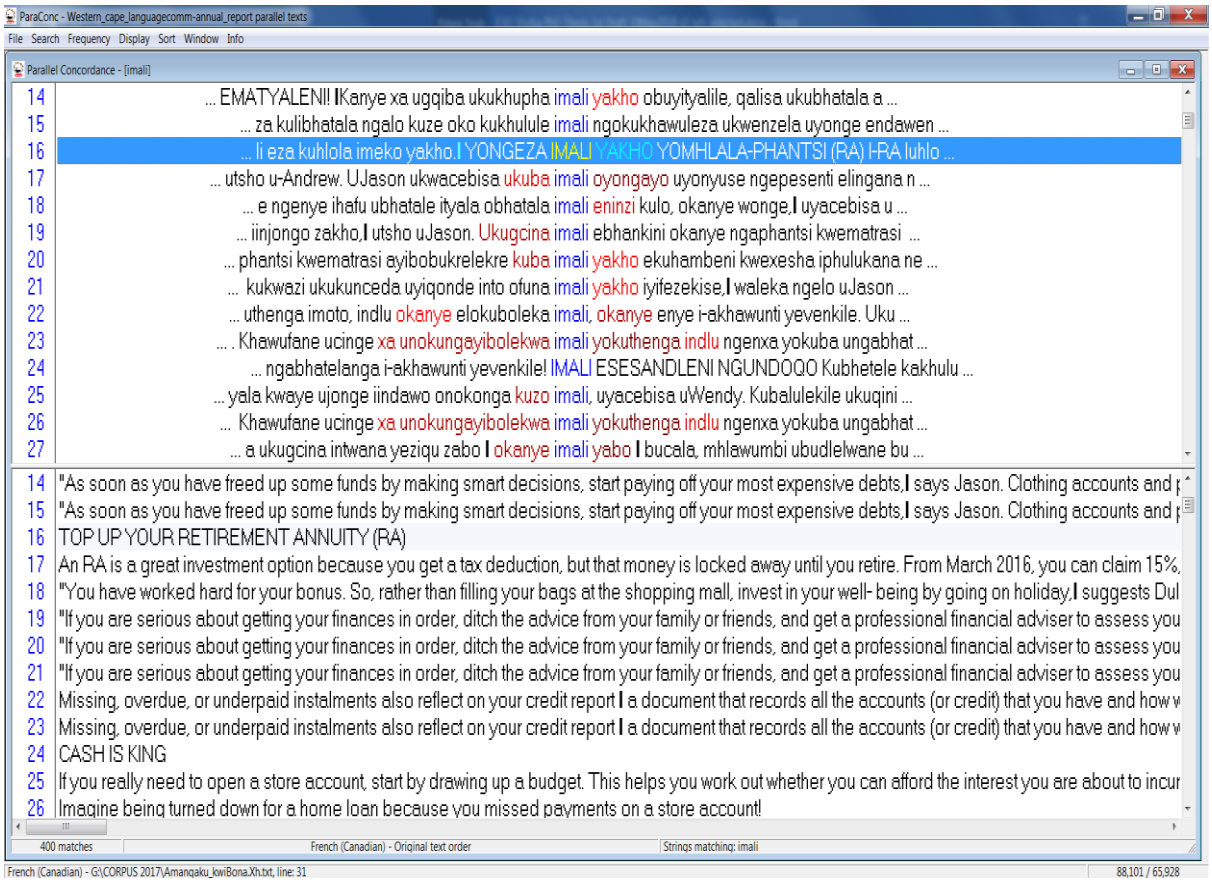


Figure 3.16: Concordance for *imali*

The English search term ‘financial’ has 500 matches whilst *imali* displays 400 matches. The concordance window displays the information as it appears in the original text order, which can be re-arranged in various ways. The following section discusses sorting of concordance results.

3.5.8 Sorting and categorising the concordance results

ParaConc can also allow one to re-arrange and sort the concordance results in order to thoroughly examine the KWIC display and concordance lines. This can be only done by selecting **Sort** from the main menu box. When you click on the **Sort** function, a sort menu box with a number of sort options pops-up is displayed in Figure 3.17. The sorting can re-arrange the concordance information alphabetically or group similar instances together, according to the query.

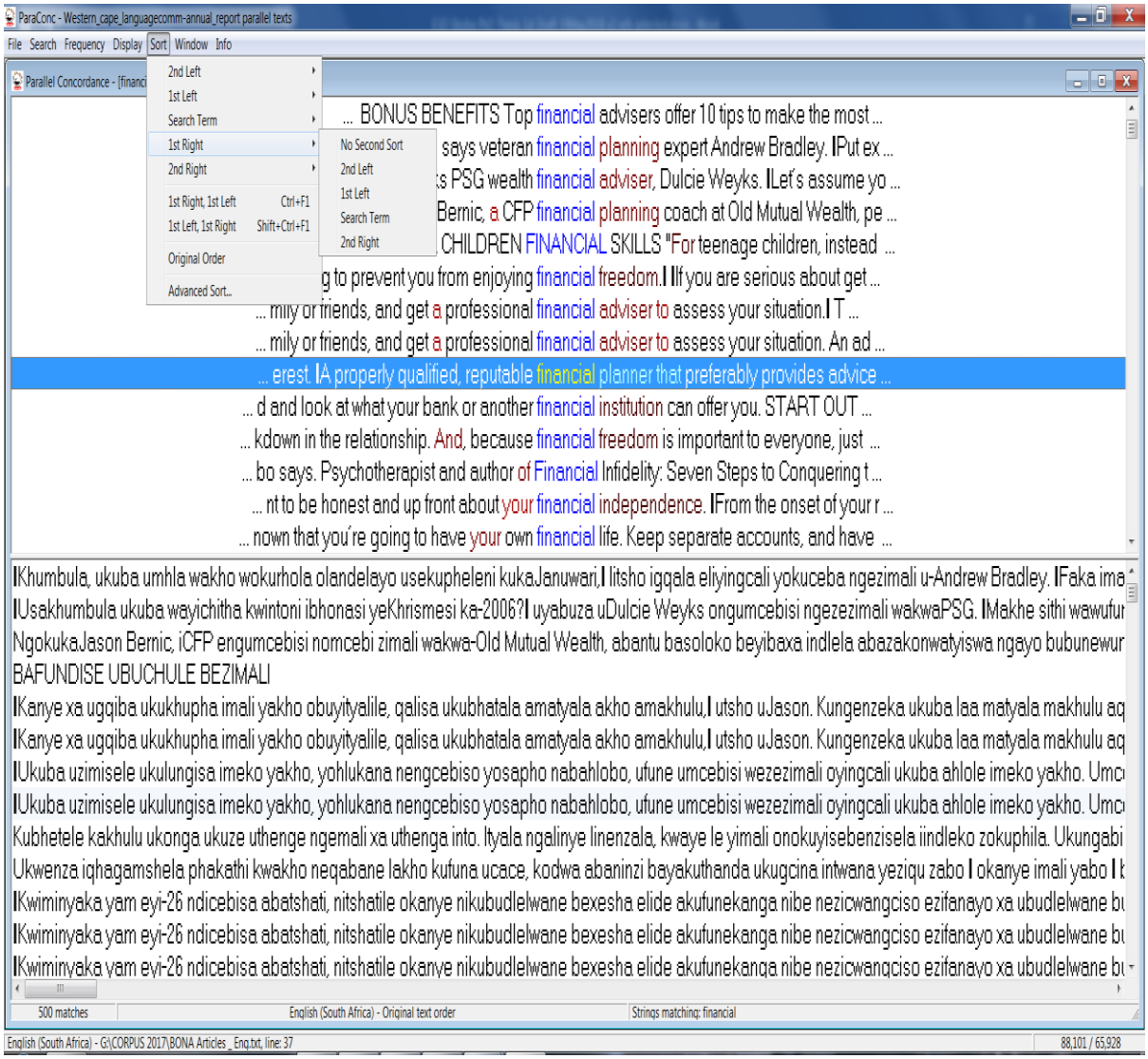


Figure 3.17: Sorting the concordance of 'financial'

The following diagram is a sorted concordance.

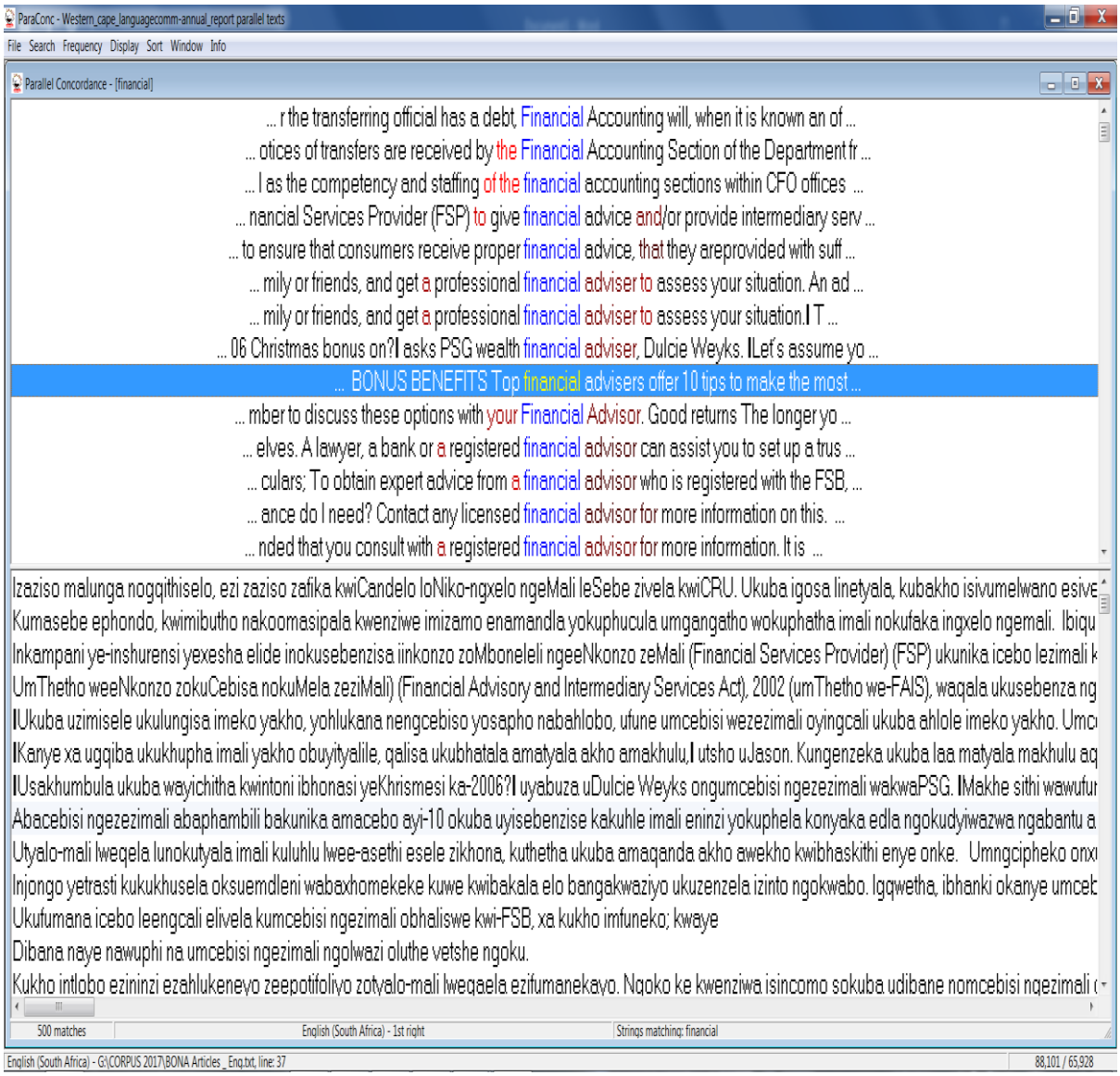


Figure 3.18: Sorted concordance results of 'financial'

The screenshot in Figure 3.18 shows the concordance results of 'financial' after clicking Sort on the menu box. In ParaConc, one can specify a primary sort by selecting 1st left and secondary one by 1st right. The researcher in the diagram above selected the **1st right** sorting order which is also described at the bottom of the results as – "**Search word 1st right'**. As seen in Figure 3.18, the original text order of the search pattern has been ordered alphabetically with the same ranges showing the same colouring. All the 1st right hand collocates of 'financial' begin with letter 'A', e.g. 'financial accounting', followed by 'financial advice', thus indicating an alphabetic ordering or sorting. Another observation is that collocates which share the same properties have

the same colour, for example, the grammatical articles for, to, etc. ‘The visual patterning displayed in the output assists in the quick scanning and analysis of the cotext’ (Barlow, 2003:33). ParaConc allows the manipulation of the various sort menu items. The parallel concordances are what makes a parallel corpus unique. In dictionary making, concordances facilitate the retrieval of possible translations of a term and how it is used in various contexts. The results were saved in the workspace for further investigation. When the concordance window is active, an option for identifying hot words is possible.

3.5.9 Hot Words

This section discusses ParaConc’s feature in which possible translations and other associated words (collocates) are suggested by the program itself – the hot words feature. These are sometimes known as keywords. The keywords are the words which occur with an unusually high frequency in a text or corpus (Bowker and Pearson, 2002:114). Hot words or key words can be searched by using the **Search** option and then clicking on **Hot words** as shown in Figure 3.19 below. To allow the program to suggest keywords, a cursor was placed in the lower half of the results window and the researcher clicked using the right mouse button; this brought up a dialogue box with ranked list of hot words (see Figure 3.19 below).

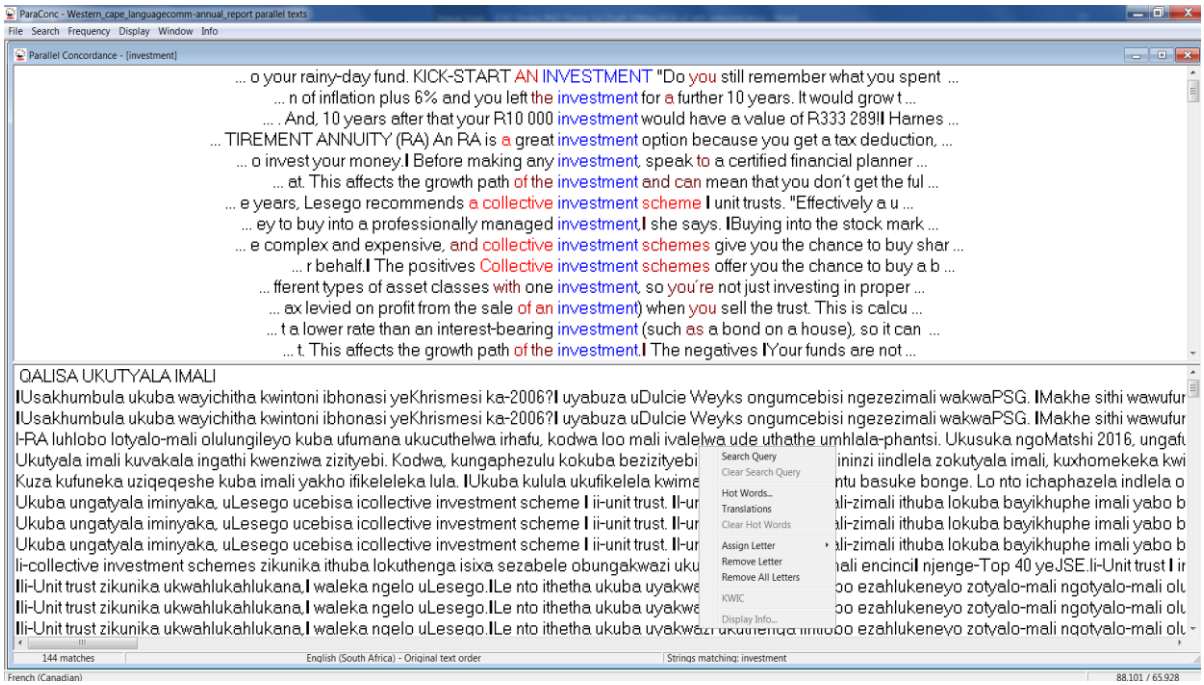


Figure 3.19: Searching for hot words

The above figure shows a command box with Hot Words command.

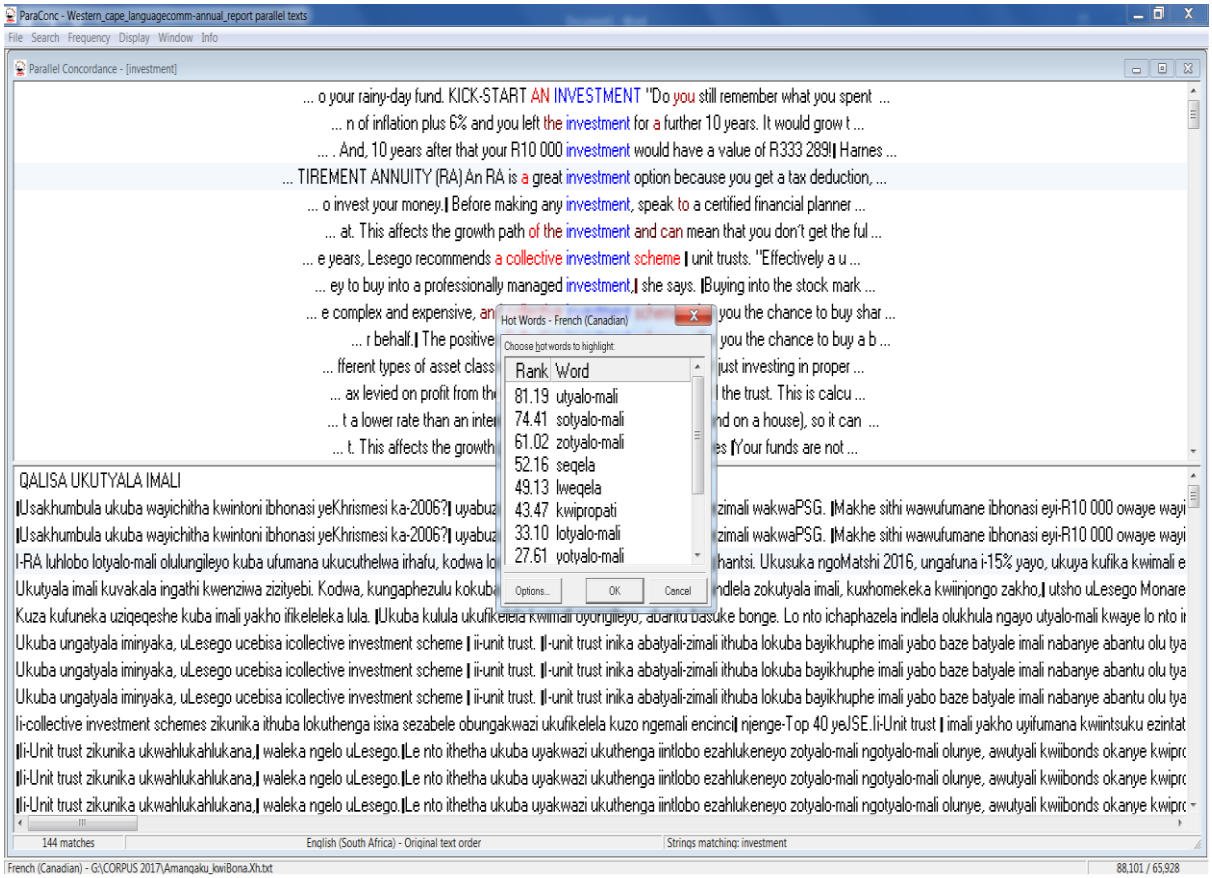


Figure 3.20: Hot words for *utyalo-mali*

Figure 3.20 illustrates the hot words associated with *utyalo-mali*. The list of words marked hot words in the small box above contains the key words suggested by ParaConc in the target language, isiXhosa. The results produced by the hot word command in an agglutinating language like isiXhosa 'enables one to search for possible translations and other associated words' (Moropa, 2005:72). The results indicate the translations of 'investment' in the EngXhPC. The highest ranked hot word is *utyalo-mali*, a translation equivalent of investment with 81.19%, while the following term is *sotyalo-mali*. It is noticed that the words are listed according to their keyness or how hot they are. The reading of the hot words in isiXhosa needs linguistic knowledge due to the morphological structure which is agglutinating. The hotwords with the stem *–tyalo-mali* appear with different prefixes as in ***lo****tyalo-mali*, ***so****tyalo-mali*, ***zo****tyalo-mali*, ***yo****tyalo-mali*. The forms prefixed to the stem *–tyalo-mali* in the list are all concords with different grammatical roles in isiXhosa. They are either possessive concords or plurals.

Words that are associated with utyalo-mali are also displayed in the hot word list. The hits that are ranked high are *lweqela* (collective) and *kwipropati* (property). This means that *utyalo-mali* is most often connected with property or group in the corpus texts. The extracted information in Figure 3.20 can be useful for a lexicographer because it tells the 'aboutness' of the EngXhPC. The role of keywords and aboutness in corpus linguistics aids in content analysis, stylistic analysis and uncovering of salient features in corpus texts (Scott, 1999). In this context 'aboutness' refers to the nature of the corpus. In addition to identification of possible translations and collocations, hot words can provide insight into the content, topic and subject of the text. This could assist a lexicographer in comparing various texts in the corpus. In dictionary making, hot words identified in Figure 3.19 above may assist in identifying the significant words, translations, synonyms and linguistic patterns.

This section has shown how concordances can be investigated to create hot word lists; the following section discusses the extraction of collocates and frequencies.

3.5.10 Extracting collocates and frequencies

One of the significant advantages of a corpus-based method is the extraction of collocates and frequencies, which has been a shortcoming of non-corpus-based methods. The following section explains how ParaConc is employed in searching for collocations.

3.5.10.1 Extracting collocates

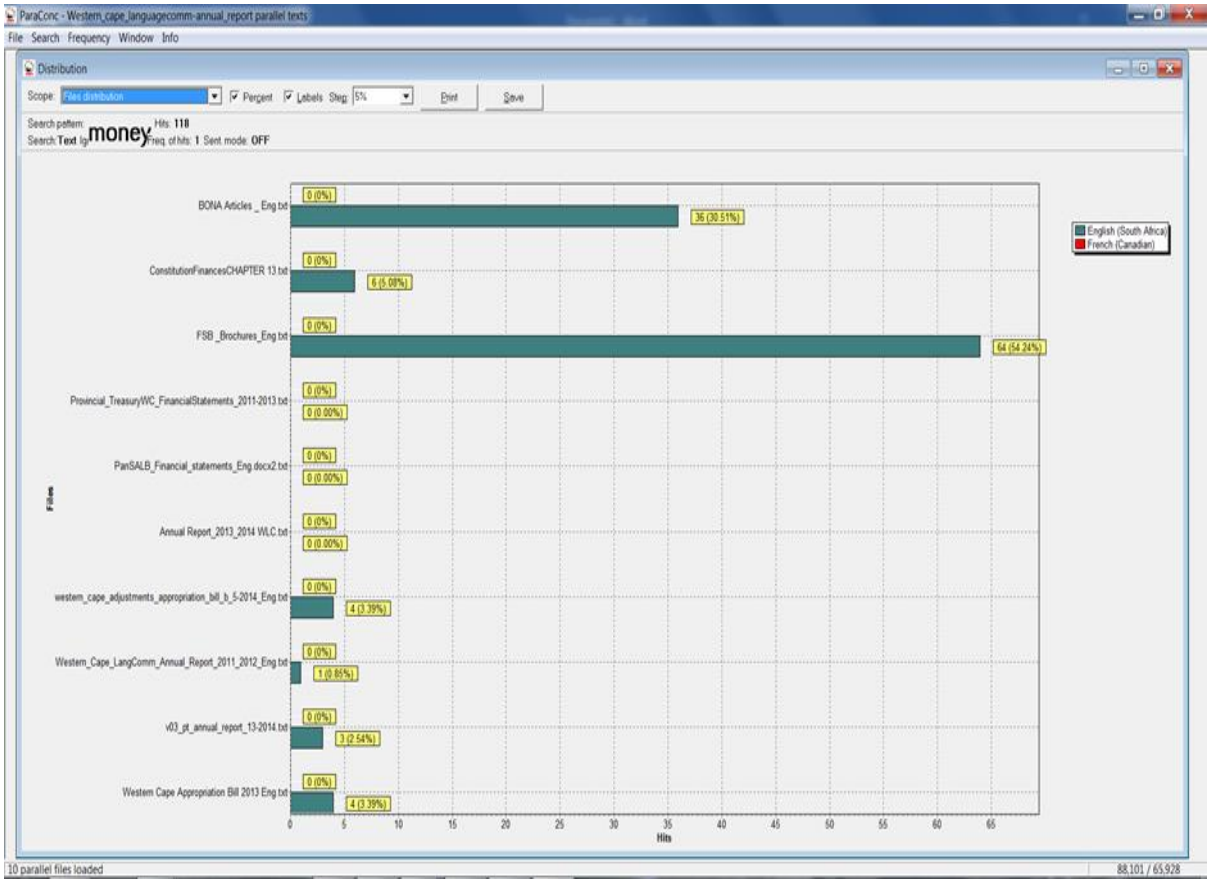
Although collocates are displayed in the concordance lines, it is also essential in this section to also look at how collocates can be queried by ParaConc. Collocates form another important type of information that can be successfully extracted from the corpus. In querying collocates one has to follow the same steps when searching for a phrase or word. Barlow (2003:104) explains that, 'collocates are tied to search terms and so there is no way to simply extract the collocations in a text'. In Figure 3.13 above, the steps for a simple search are displayed in the search menu box. To focus on the extraction of collocates one has to enter a group of words in the search menu, click on **Advance collocation** and then select the **Span** option. The examination of collocates according to Bennet (2010) provides a deeper understanding of the meaning and the use of the word. The denotation of words and their usage is important information to be presented in a domain-specific dictionary.

3.5.10.2 Collocate frequencies

'ParaConc furnishes a variety of frequency statistics, but the two main kinds are corpus frequency and collocate frequency' (Barlow, 2008:65). Collocate frequency provides information about the number of occurrences of collocates of a search term. To generate collocates frequency information on 'financial' and *yemali* respectively, the frequency menu was opened and clicked on. Collocate Frequency Data displayed collocates of the two search terms in both the descending order and the frequency order. Collocates frequencies can be very useful in showing up various kinds of information that are of the greatest interest to the current study such as synonyms and meaning identification.

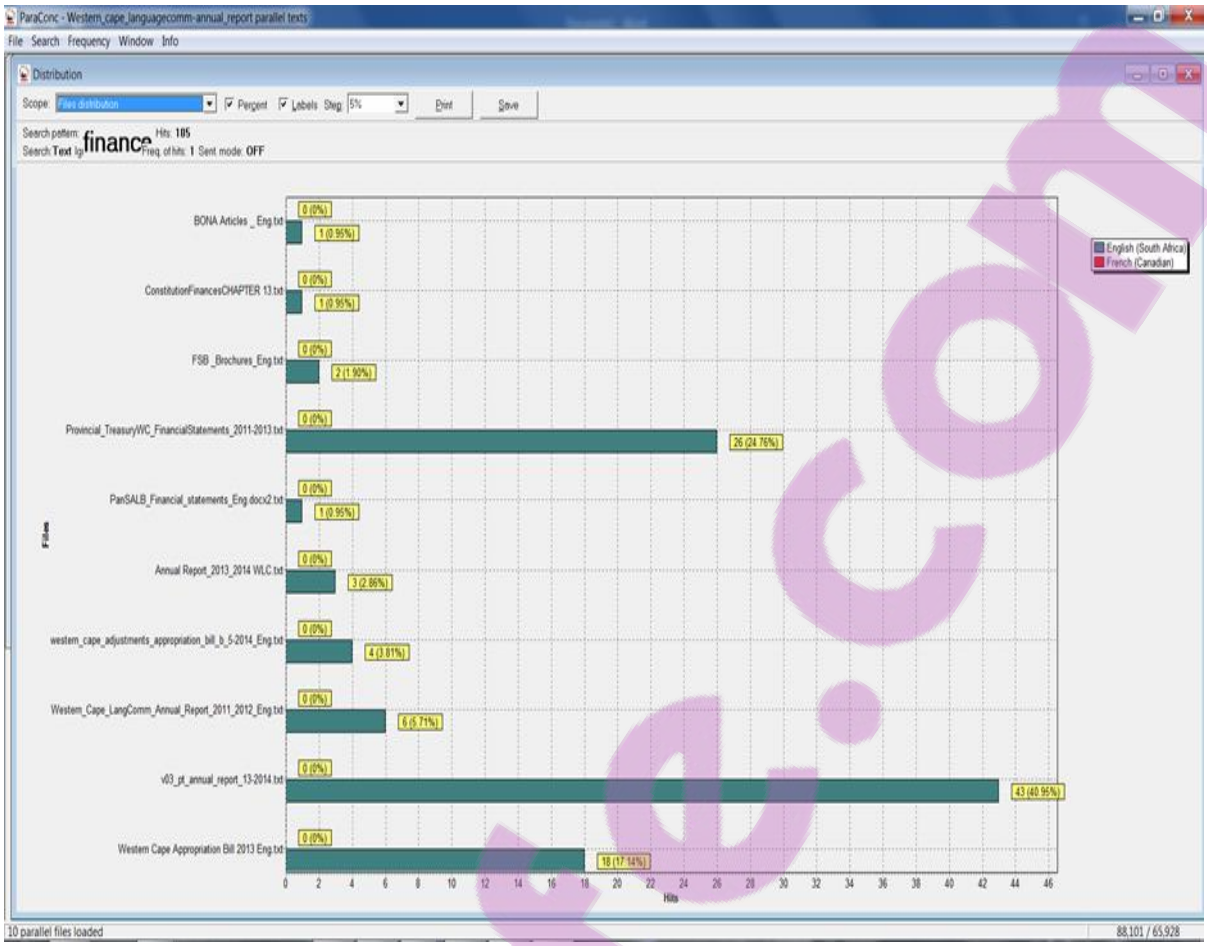
3.5.11 Distribution

ParaConc offers a feature that allows the researcher to investigate how words are distributed in a corpus or in a particular file. The distribution function can only be performed when the concordance results window is active. To obtain the distribution information, one has to click on **Display** on the main menu. A drop-down box with a list of options appears, and the **Distribution** command is then selected by clicking on it. The output is displayed in the form of a graph as shown below.



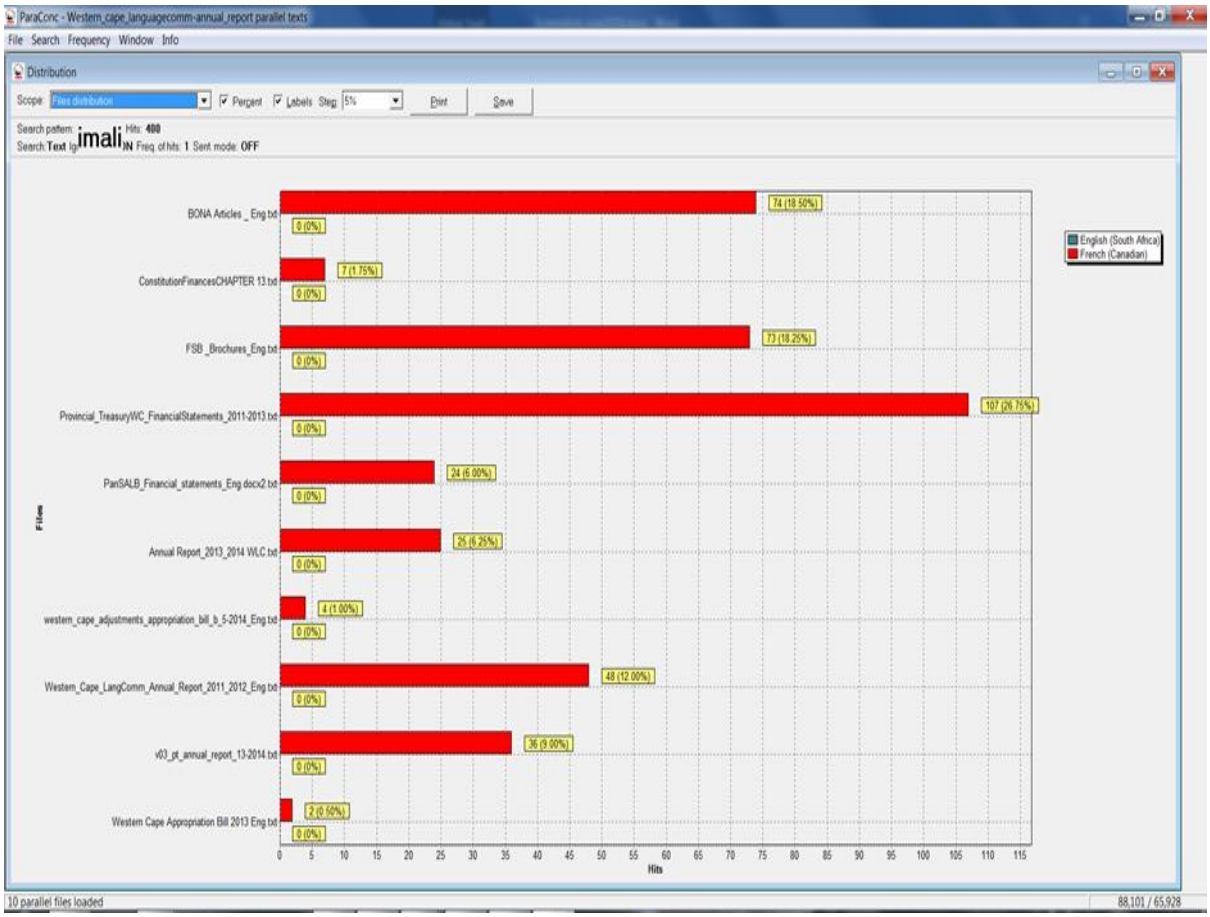
Graph 3.1: Distribution of money*/imali*

Graph 3.1 illustrates how the word ‘money’ is distributed in the English and isiXhosa parallel texts.



Graph 3.2: File distribution of finance/imali

Graphs 3.1 and 3.2 illustrate how the words ‘finance and money’ are spread in the source texts of the parallel corpus. The word ‘money’ shows 118 hits whilst ‘finance’ has 105 hits. The bar indicates that the word does appear in the file; when the bar is non-existent, such a word is not recorded in the file. For example, in Graph 3.1 texts 4, 5 and 6 have 0 records concerning ‘money’ and are shown in yellow with 0%. The word ‘finance’ is distributed across all the texts in the corpus with, for example, text 4 showing 24.76%.



Graph 3.3: The distribution of *imali*

The graph above illustrates the distribution of the target language word *imali*. In contrast with the source terms ‘finance/money’, *imali* records 400 which is quite high. The reason is attributed to the fact that both terms are translated as *imali* and regarded as synonyms in isiXhosa: ‘money’ and *imali*. The word *imali* occurs in all corpus files but the percentage of its use in the annual reports is very low whilst in the articles and brochures for the public it is high. In Figure 3.3 *imali* is better distributed than its counterparts.

3.5.12 Summary

This section first discussed and justified the research methodology. In order to address the aims of the study, a parallel corpus of English and isiXhosa was designed. The corpus design criteria were considered before the actual compilation and analysis of the corpus. In considering the said criteria, the researcher had to make decisions on the identification of texts to be included in the corpus, the size of the corpus and copyright or ethical issues. In compiling the EngXhPC, ParaConc, a parallel concordancer was used to load the corpus files and thereafter used to perform various corpus query functions such as frequency lists, collocation, equivalents. The following section provides an outline of how ParaConc will be used in the analysis chapter to extract bilingual financial terminology from the English-isiXhosa Parallel Corpus.

3.6 Exploiting ParaConc to extract bilingual terminology

The aim of this section is to demonstrate how ParaConc will be used in the analysis chapter to achieve the second aim: the extraction of bilingual terminology that can be used to create a specialised bilingual dictionary. In section 3.5 above, the EngXhPC composed of parallel texts in finance was created. The parallel concordancer, ParaConc, was selected to perform different tasks in order to conduct a quantitave and qualitative analysis of the parallel corpus. The manipulation of EngXhPC by means of ParaConc is indicative that corpus analysis software is able to carry the counting, sorting and presentation of language features (Baker, 2010:103). As a concordancer, ParaConc's analysis of this parallel corpus produced bilingual wordlists, frequencies, concordances and other statistical information. All the functionalities of ParaConc can be used in extracting information for dictionary making.

3.6.1 Corpus evidence applicable to a bilingual dictionary

Different scholars identify three stages of the dictionary-making process where corpus data can be used: (i) at the establishment of the lemma list, (ii) entry preparation and (iii) writing or editing stage (cf. Atkins and Rundell 2008; Landau, 1984; Roberts, 1996). An ideal bilingual dictionary presents at least three kinds of information about words, namely, i) the form, ii) meaning and iii) context. The information on form includes spelling, variants and inflections, while meaning includes translation equivalents, synonyms; context refers to the real usage of a word in a textual environment. The subsequent section indicates how ParaConc has been used in extracting bilingual information from EngXhPC.

3.6.1.1 Lemma lists

The EngXhPC can be used in generating developing a headword/lemma list. Lemma lists are obtained from the frequency list either in the order of frequency or of the alphabet (see Figure 3.10 and 3.11). The frequency lists generated in Figure 3.11 indicate that the corpus is for special purposes because the most frequent words are finance-related terms. For example, the word 'financial' in Figure 3.12 is highly frequent with 1042 counts, while *imali* has 400 counts. This means that these two words are possible lemma candidates in a dictionary of financial terms. 'A lemmatised frequency list can be obtained from word frequency studies on a lemmatised electronic corpus' (De Schryver and Prinsloo, 2000:297). Word lists in frequency order cannot be developed from other parallel texts. Compiling a lemma list is the first step in dictionary making and in the electronic era frequency based dictionaries are becoming a trend. The EngXhPC is aligned in order to enable various parallel searches. A word list according to Bowker and Pearson (2002:119) can be used for a number of purposes:

- To familiarise oneself with the terminology in the corpus
- To identify words which are related to each other
- To observe typical co-occurrence pattern in word clusters.

The frequency lists generated from EngXhPC not only assist in the inclusion decisions but can also be of use in providing translation equivalents in order of frequency. Further information on frequency lists is discussed in section 4 of the analysis chapter.

3.6.1.2 Orthographic information

The generated frequency lists, for example, display the orthography of words in both English and isiXhosa (see Figures 3.11 and 3.12): 'Orthography is the writing system in standard everyday use, which consequently attracts most studies' (Crystal, 1987:204). A standard orthography provides a uniform way of writing and spelling in a language. Information on how a word is written or spelt is necessary in any dictionary. Determining the spelling of the headwords and their translations from the parallel corpora is central in bilingual lexicography. Besides the frequency lists, the concordance results offer a clue to the spelling rules followed in the parallel texts. The equivalent translations of the English word 'financial year' have two orthographic representations: i) *unyaka-mali,* and ii) *unyakamali,* whilst in English two spelling variations*:* 'advisor and adviser' are easily identifiable in the sorted concordance results in Figure 3.17. The two nouns in the source language are derivatives of the verb 'advise'. In the English concordances it always appears in the company of the adjective financial as in 'financial adviser'. The parallel corpus shows that some translators have spelt it as 'advisor' according US English spelling. In English the two spellings are acceptable and in a dictionary they should be included. The use and non-use of a hyphen between vowels is also displayed in the EngXhPC and is another example of inconsistencies in isiXhosa orthography. For example, the most common translation equivalents of the SL word 'asset' extracted by ParaConc are *ii-asethi, iaseti* and *iiasethi*. This is due to changes in the orthographic rules. In the revised orthography and spelling rules of isiXhosa published in 2008, the hyphen is commonly used in compounds whereas in the previous standards a hyphen in compounds was optional or not used.

The spelling variations investigated by the software are indicative of different styles of writing by the various translators, which require further investigations by the compiler. The retrieval of orthographic information is very useful when deciding on which orthography to choose for a specific dictionary. For a lexicographer to be able to decide on which word form to include in the dictionary, other sources like orthography and spelling rules need to be consulted in order to establish the standard form. This shows that whilst the corpus-based method of creating dictionaries is quick and more accurate, it is not exhaustive, as there is a need to go beyond the dictionary to make other lexicographic decisions.

3.6.1.3 Identification of possible translations

A bilingual dictionary presents information in two languages; thus, it is important to translate words from the source language to the target language. The process of translating words is usually time consuming, but this process is simplified in the parallel corpus because all the words have one or more translations, which is a bonus. The aligned parallel texts of the source and the target language create an enabling platform to spot translation equivalents. The semi-automatic aligner of ParaConc links and creates connections between L1 and L2 sentences. The possible translation equivalents are displayed in context. In cases where some words are in isolation, ParaConc generates corresponding equivalents such as headings and definitions. The following examples are taken from the aligned texts illustrated in Figure 3.9 above:

Table 3.2: Identifying translation equivalents

Source Language	**Target Language**
1. Values	*Iimpawu zentsulungeko*
2. Accountability	*Ukuba noxanduva lokuphendula*
3. Competence	*Isakhono*
4. Care and	*Inkathalo kunye*
5. Responsiveness	*Nokuphakama msinya*

The table above displays translation equivalents after search query option was used. It is important to note that a parallel corpus generally provides more than one translation equivalent, especially in cases where different translators translated the texts. Lexicographers therefore have the difficult tasks of choosing the most appropriate of the available equivalents. The next diagram shows how the term amount is translated in isiXhosa.

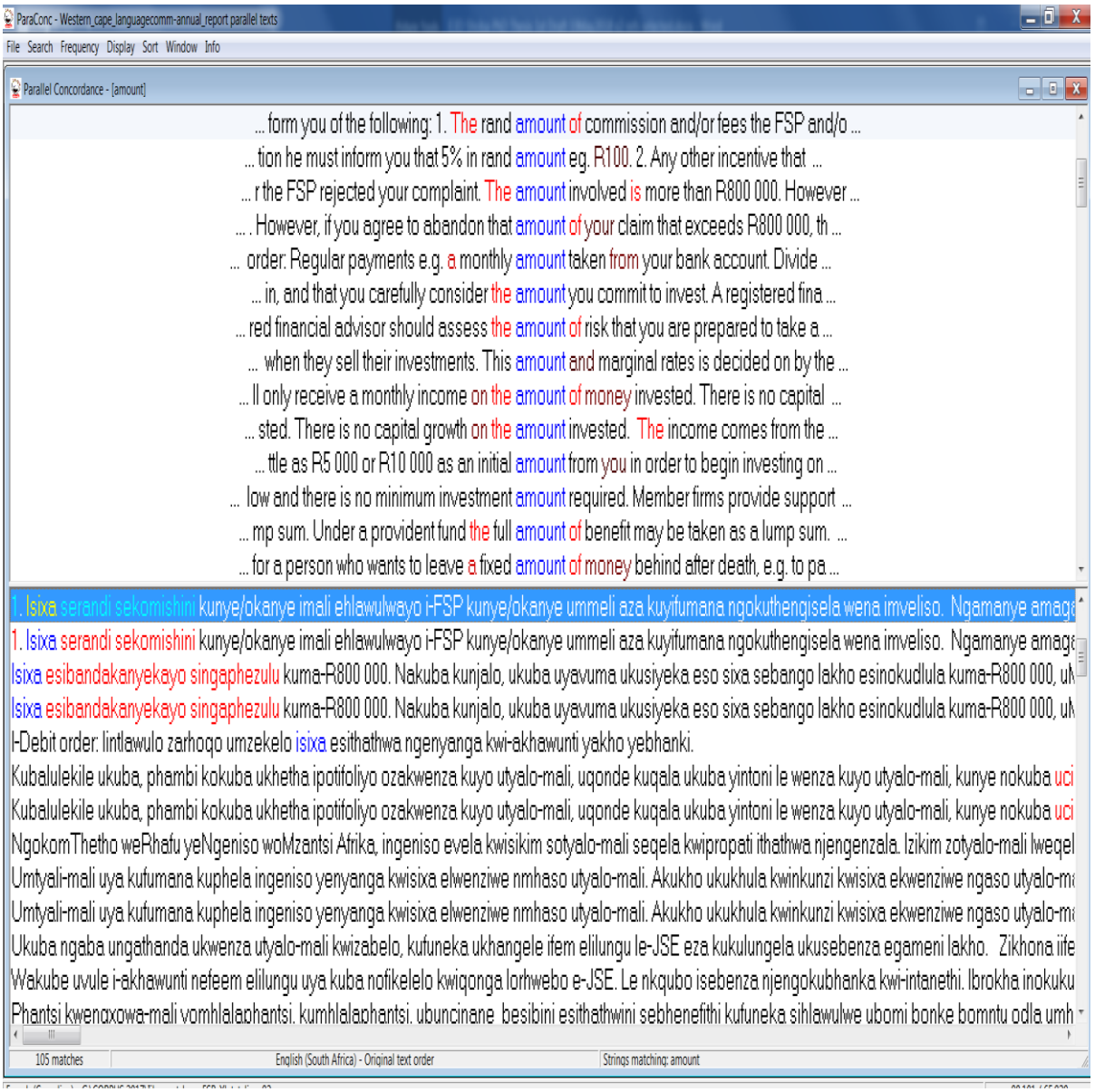


Figure 3.21: Translation of amount into a target equivalent *isixa*

Figure 3.21 illustrates how ParaCanc displays the translation equivalent of the source word 'amount'. The upper window of the concordance displays the search word in the source language. The results indicate *isixa* as the translation of amount. In examining the concordance lines, the equivalent is written as *isixa mali,* a non-hyphenated compound whilst in the other examples, amount is translated as *isixa* with the omission of the second stem –*mali*. The search feature is the most convenient and the quickest option for identifying possible translations; however, this option is not as quick as the search function of ParaConc which is fully supported in the study because of its efficiency.

3.6.1.4 Retrieving usage examples

The inclusion of example sentences in most dictionary typologies is very important because it aids in meaning clarification especially when the former are carefully selected. This process is simplified in parallel corpora because words are always presented in context. Usage examples are searched from the concordance information, and give the exact behaviour of a word in use. In the concordances of risk/*umngcipheko* the following example sentences can accompany the translation equivalents in order to provide more semantic information on a specialist term:

1. ST: Minimise the **risk** of under-spending.
   TT: *Cutha* ***umngcipheko*** *wenkcitho encinci*.
2. ST: Effectiveness and efficiency of **risk management**.
   TT: *Impumelelo nocikizeko* ***loLawulo Lwemingcipheko***
3. ST: **Enteprise Risk Management** (ERM) is centralised provincially within Department of the Premier.
4. TT: **U*Lawulo Lwemingcipheko yeZiko*** *(ERM) libekwe ndaweni nye ngokwalapha kwiphondo eli kwiSebe leNkulumbuso yePhondo.*

The example sentences are taken from the KWIC of concordance view of *umngcipheko*/risk, the upper window and the lower window. This would provide relief for the lexicographer because these examples are context-dependent. The corpus benefits are attested by McEnery and Xiao (2016:443), ‘Before corpora made inroads in bilingual dictionary making, the examples in such dictionaries were either invented introspectively by lexicographers or cited from famous literary works’. This is indicative of the power of corpus data: that it is more efficient and laden with useful data compared to traditional methods where intuition is often used to make decisions. As intimated, intuitively crafted examples are usually subjective and sometimes not appropriate. The analysis of KWIC information by the researcher is a proof of the benefits of this resource. The corpus analysis tool, ParaConc, only generated quantitative information in the KWIC format, which requires human analysis. The two sentences can be used in a dictionary to further enhance meaning and indicate how the search word is used in context:

3.6.1.5 Identifying Multiword Units

A multiword unit can be simply defined as any lexical unit that is composed of more than one word in order to create a new word. The generated concordance lines display co-occurrences, which is how multiword lexical units are recognised. These words are found in large quantities in a naturally occurring language in corpora and are recognisable through corpus analysis tools. The specialised parallel texts are marked by a high frequency of phrasal verbs and compound nouns. When looking at both frequency and alphabetic lists, ParaConc only provides single words, leaving out word forms that are constituted by more than units. Concordances lines easily display the multiword units prevalent in the parallel EngXhPC both in English and isiXhosa, for example:

i) conditional allocation = *ulwabiwo olunemiqathango*
ii) municipal financial resources = *imithombo yezemali kamasipala*
iii) conditional grant = *isabelo esinemiqathango*
iv) pension fund = *ingxowa-mali yomhlalaphantsi*.

The specialist texts are characterised by compounds with two or more words; this information is found from the reading and analysis of the concordance lines. Compounds are common in specialised fields in the Xhosa language and are an indication of the creativity of Xhosa translators. These are important entries in specialised bilingual dictionaries and the **Sort** feature makes it possible and easier to identify these.

3.6.1.6 Grammatical and structural analysis of headwords

ParaConc does not generally mark parts of speech or word categories; nonetheless an analysis of the context in the concordance was able to tell whether a word is used as a noun or verb. Examples of verbs and nouns are recorded in Table 3.2 below.

Table 3.3: Nouns and verbs in the parallel corpus

Verbs in SL	**TL equivalent**	**Nouns in SL**	**TL equivalent**
Accrued	*eyongezelelwayo*	virement	*Ivayamenti*
Audit	*Ukuphicotha*	rates	*Iirhafu*
Pay	*Ukuhlawula*	allocation	*Ulwabiwo*
Fixed	*yabekwa bucala*	claims	*Amabango*
Transfer	*Ukuwelisela*	grant	*Isibonelelo*

The table above exemplifies the content nouns and verbs that are in the parallel corpus. This kind of information may help during lemmatisation. The inflexions in verbs can tell the tense, such as accrued in example 1. Corpus analysis can also recognise the common structures of headwords in sentences. It should be noted that a corpus rarely yields single words, so words are always part of syntax. The following sentences in isiXhosa, for example, illustrate the structural behaviour of the headword *unyaka-mali* when preceded by ingxelo:

1. *Ingxelo **yo**nyaka-mali ophele ngomhla wama-31 kweyoKwindla* 2012 (An annual financial report for the year ended 31 March 2012).
2. *Ingxelo **ku**nyaka-mali odlulieyo* (A report in the previous financial year).

In the examples above the node *-nyaka-mali* may be preceded by either a locative or possessive formative. This indicates that the corpus can also help in identifying the correct concords by analysing words in context.

3.6.1.7 Identifying collocations

Collocations are only identifiable through the evaluation of concordance information and are not easy to pick up manually. These comprise words that usually co-occur or go together in a pattern. ParaConc's ability to retrieve and display concordance lines makes it straightforward for the corpus analyst to study the collocational profile of a node word or lemma. Kenny (2001:81), describes a collocation as 'a syntagmatic relationship between at least two lexical items'. In Figure 3.14, the search function of the word 'financial' displayed the search term in the centre with other words on either side, differentiated by different colours. This is indicative of the internal structure of the collocation which is made up of a 'node and collocates'. 'The node is the item whose lexical behaviour is under investigation; the collocates those items that occur in its environment' (Kenny, 2001:87). In order to discover the collocates of the node 'financial', human intervention completed the semi-automatic extraction by reading the concordance lines retrieved by ParaConc. In the window displayed in Figure 3.14, the adjective 'financial' co-occurs with a number of nouns such as: **annual, year, assets, resources, management** and so on. Reading from the left to the right of the search word compound, noun phrases are picked up such as:

i) annual financial year
ii) financial assets
iii) financial resource and financial management.

The information on collocation is easily obtained from the corpus and is significant in corpus-based dictionaries. The EngXhPC consists of texts that are translated by various translators for different genres dealing with finance or money matters, employing various translation strategies or term creation processes. These will be analysed in the next chapter.

3.6.1.8 Summary

This section demonstrated how Paraconc was used to extract bilingual terminology for dictionary making. Through ParaConc, the EngXhPC was queried with a view to extracting bilingual information that could contribute to the making of a specialised dictionary of finance. The features that were explored are the alignment, word lister, frequency, concordances and identification of collocations.

3.7 Conclusion

The aim of this chapter was to outline the methodology and analytical frameworks followed by the researcher in achieving the research aims stated in Chapter 1. The corpus-based method selected in this study is characterised by quantitative and qualitative approaches that were complementary in the data collection and analysis stages of this research. The following chapter will deal with the analysis, interpretation and discussion of the findings.

CHAPTER 4

Findings and interpretations

4.1 Introduction

Chapter 3 discussed the research design, methodological procedures, ethical issues, data collection procedures, creation of a parallel corpus and evidence of how the corpus works. In order to achieve the aims of the study stated in the introduction chapter, a specialised corpus of English source texts (ST) and isiXhosa target texts (TT) was created. ParaConc was selected as an analysis tool.

The aim of this chapter is to present and analyse the collected data in order to show how ParaConc can be used to create a bilingual specialised dictionary, and also demonstrate how corpus evidence and data can significantly contribute to the compilation of bilingual specialised lexicography in English and isiXhosa. The data categories that will be uncovered from the EngXhPC by means of ParaConc will be presented, analysed and interpreted under the following main themes: word lists, translation equivalents, concordance information, synonyms, multiword units and collocations.

In interpreting and discussing the findings, the following reference works: *Oxford English-Xhosa Dictionary (*1985*), the GDX (Volume 1, 2, 3: (*1989-2006*), Oxford South African Concise Dictionary (2002, Oxford Finance and Banking Dictionary* (2014) and the *Orthography and Spelling Rules for isiXhosa* (1972 and 2008) were consulted. The following factors informed the selection: contemporariness, being subject specific, locality and popularity in the two languages and dealing with finance as the subject of this study. During the analysis stage, the *Oxford English-Xhosa Dictionary* was used to check the meanings of work in relation to how the English source words were actually translated by translators. This reference work is the most commonly used descriptive bilingual dictionary targeted at both English and isiXhosa speakers. It is appropriate because of the entries that cover everyday, technical and subject-specific English. *The GDX* was also selected to look up various types of information on target language terms extracted from EngXhPC. It was selected because it is a comprehensive trilingual dictionary (IsiXhosa, English and Afrikaans) with isiXhosa

entry words. The first column in isiXhosa functions as a monolingual dictionary with detailed information on each lemma. For analysis of orthography of isiXhosa words, *GDX* and the revised orthography book (2008) were compared. The *Spelling and Orthography Rules* booklet contains the latest convention and spelling rules of isiXhosa that are helpful for interpretation purposes.

The *New Oxford South African Concise Dictionary* (2010) is 'the most up-to-date South African English dictionary available' (see the cover page). As a general dictionary that covers all the English words including some commonly used specialist terms, it was selected in order to search for meanings of words used in the SL texts. It is important to state that presently there is no English-isiXhosa bilingual dictionary of financial terms and that the dictionaries that are used in this study are mostly general and not exhaustive. For specific and specialised meanings of financial source terms, the *Oxford Dictionary of Finance and Banking* (2014) was selected. This monolingual specialised dictionary is mostly used by translators to look up meanings of terms during source text analysis. In this study it was used to verify definitions of financial terms and whether these terms were translated correctly in the target texts.

In Chapter 2, section 2.7.1.2 the distinction between a monolingual and bilingual dictionary was drawn. As stated, bilingual dictionaries contain two languages and are the most often consulted dictionaries in multilingual societies, referred to by a variety of users for different functions such as reading, writing, translation and comprehending difficult and specialised concepts. For this reason, they are described as polyfunctional. Gouws (1996:14) contends that, as a polyfunctional instrument, 'a bilingual dictionary by implication presents more information than just translation equivalents'. This statement is very relevant to a specialised bilingual dictionary in a language with scarce lexicographic resources. Apart from the headwords and the target equivalents, essential components of a specialised bilingual dictionary entry include amongst other details: grammatical information, orthography, semantic information, synonyms, contextual meaning, abbreviations/acronyms, collocations and other multiword units; these will be explored in this chapter. The following section provides the interpretation of the findings under the following main headings: headword lists, translation equivalents and translations, orthography, synonyms, concordancing, contextual meaning, collocations and other multiword units.

4.2 Extracting headwords

A bilingual dictionary is comprised of headwords and their equivalents in the target language. Hartmann and James (1998:67) define the headword as ‘the form of a word or phrase which is chosen for the lemma, the position in the dictionary structure where the entry starts’. Headwords are entry words that are usually presented in an alphabetic manner in the dictionary. The parallel corpus can be used to identify headwords through the alphabetic list which is presented below.

4.2.1 Alphabetical word list

In a parallel corpus, all the words can be arranged according to the first letter of the alphabet in order to produce an alphabetical list. The alphabetisation or sorting of words is automated by ParaConc, a process which takes seconds, thus confirming the efficiency and speed of the corpus software when compared to the traditional method. In the traditional method of dictionary making, as noted, arranging words in alphabetical order is done manually and is a tedious process that often results in the omission of key words or inclusion of unnecessary data. However, in a parallel corpus, all the words are accounted for as shown in Table 4.1 below which displays alphabetic lists generated by ParaConc in English and isiXhosa (French Canadian in the EngXhPC constructed for the current study).

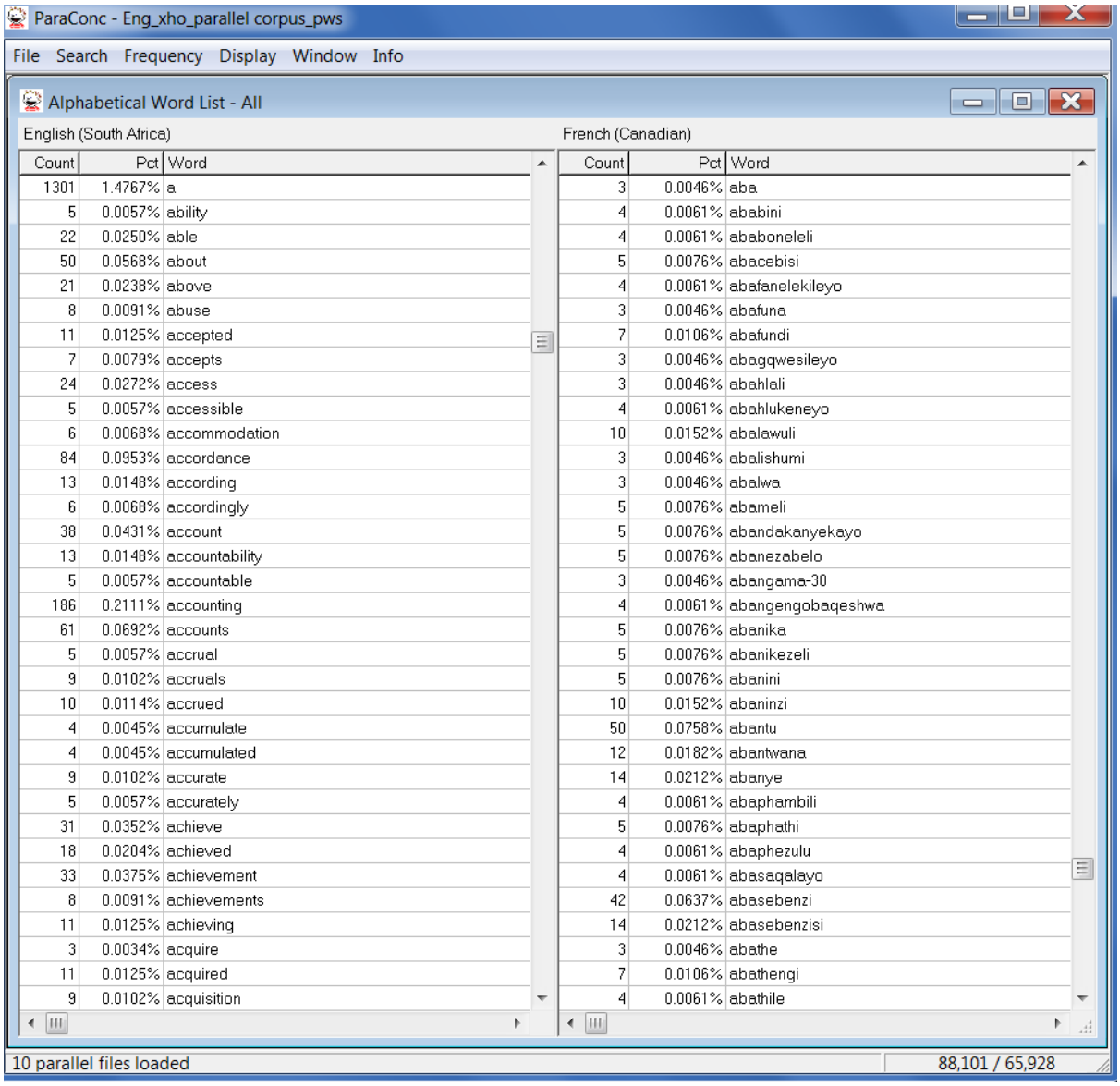


Figure 4.1: English-isiXhosa alphabetical list extracted from EngXhPC

The diagram above is an alphabetic list representing all the words that begin with the letter ‘A’. The alphabetic word lists from the EngXhPC are generated in both English as the source language (SL) and isiXhosa as the target language (TL). The two lists display all the word forms that are contained in the parallel corpus, that is, non-financial and financial words with their frequency counts. In the English alphabetic list, the word in the first position is the grammatical word, the article ‘a’, whilst in the isiXhosa list the leading word is the demonstrative *aba.* From the alphabetic lists in Figure 4.1, a lemma list can be drawn up according to the type of dictionary that is being created. Table 4.1 displays the lemma list in the source language and from it, financial terms can be drawn as exemplified below.

Table 4.1: Alphabetic word lists generated from the English-isiXhosa Parallel Corpus

Count	**Word**
5	Able
40	About
16	Above
3	Acceptable
5	Accepted
5	Accepts
30	**Access**
5	Accessible
4	Accommodation
49	Accordance
10	According
3	**Account**
15	**Accountability**
5	**Accountable**
3	**Accountants**
129	**Accounting**
45	**Accounts**
5	**Accruals**

The list in Table 4.1. illustrates the first 20 words automatically arranged by ParaConc in the EngXhPC. Because the source language of the parallel corpus under investigation is English, the researcher saw it fit to analyse the SL list. All the highlighted words (in bold) are financial terms that are used in the banking and finance

fields, among others. The words ‘account’ and ‘accounts’ are arranged according to their alphabetic order or positions, demonstrating both the singular and the plural form. In the *Oxford Dictionary of Finance and Banking* (2014:3) the term ‘account’ has at least six senses or definitions. The first definition refers to ‘a statement of indebtedness from one person to another; an invoice’. In the specialised dictionary, the very first definition relates to a field of finance whilst in the second general-purpose one the first definition is more general. On the other hand, in *the Oxford South African Concise Dictionary,* the first two definitions read as follows: n. **1** a description of an event or experience. **2** a record of financial expenditure and receipts > a bill taking the form of such a record. This shows that each dictionary is designed to serve a specific audience; therefore, the information from this specialist parallel corpus will help those users that are solely interested in understanding finance terms. Based on the definitions from the dictionary, the term account can be selected and entered as a headword in the specialised bilingual dictionary. In a bilingual dictionary, it is important to classify terms according to whether they are singular or plural, and the parallel corpus presents terms in different forms as shown by the term account (singular) and accounts (plural). Amongst the list of words highlighted in this alphabetic list, ‘accounting and accountability’ will be analysed to show their importance as headwords.

In terms of alphabetic ordering, ‘accounting’ ranks in the list with the highest word count of 129, indicative of its frequent occurrence in the EngXhPC. ‘Accounting is a noun that refers to the action of keeping financial accounts’ (*Oxford South African Concise Dictionary*, 2006:7). In the *English-Xhosa Dictionary* (1985:5) accounting is entered as a sub-lemma under the headword ‘account’ and translated as *ucalulo-mali, ucwangciso-mali.* The presence of words and their possible translations makes the process of dictionary making easier; hence the parallel corpus is advocated for as a resource for lexicographers.

The word ‘accountability’ is a financial term that is commonly used in financial management and governance. In Table 4.1, ‘accountability’ is amongst the highlighted terms with a word count of 15. In general language it may mean to account for one’s decisions or responsibilities. In a language for special purposes (LSP) one accounts for money and assets. ‘Accountability’ is lemmatised in each accounting, finance and banking dictionary. In the *Oxford Dictionary of Finance and Banking* (2014:3), for

example, accountability is ‘an obligation to give account’. Contrary to a general-purpose dictionary, the entry in this dictionary provides more information on this word. The annual reports, for example, are the annual accounts where public officials are reporting internally and externally to the stakeholders. The extraction of this word is indicative of its use in real language in the finance domain; therefore, it is likely to be searched for in a dictionary of finance.

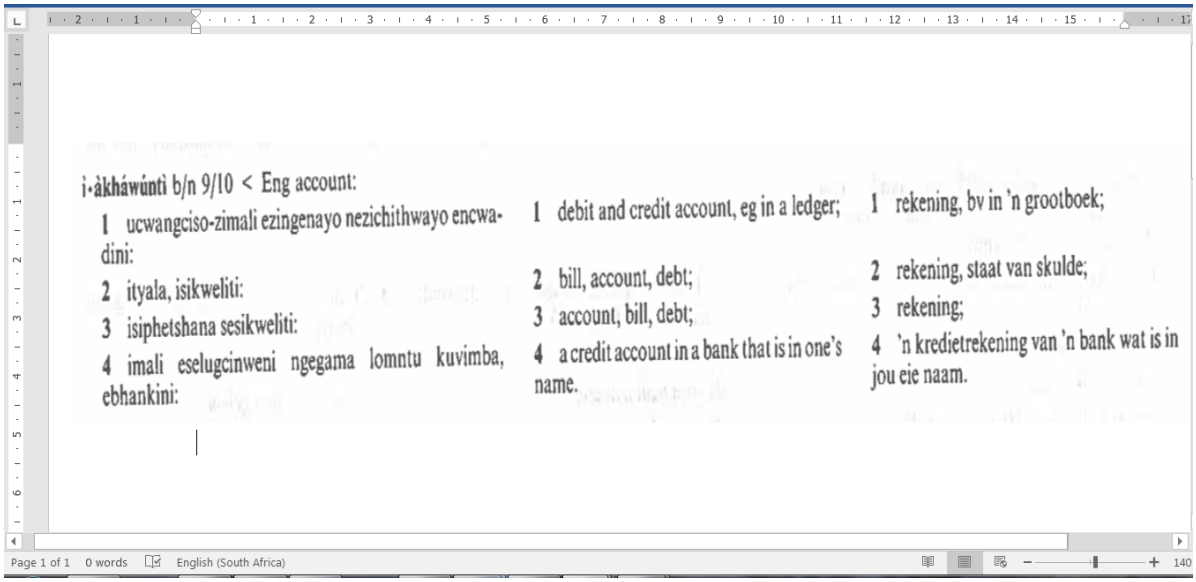
A possible term list with translation equivalents from the EngXhPC is extractable from the above comprehensive lemma lists for inclusion in a bilingual specialised dictionary (cf. Table 4.2 below).

Table 4.2 Candidate term list developed from the EngXhPC

English SL finance term	**IsiXhosa Target equivalent**
Account	*Iakhawunti*
Accountability	*uxanduva lokuphendula*
Accountable	*ukukwazi ukuphendula*
Accountants	*abacwangcisi-zimali/ababali-zimali*
Accounts	*ii-akhawunti, iiakhawunti*
Accounting	*ucwangciso-mali, ubalo-mali*
Accruals	*iimali ezongezelelweyo*

From the above discussion, it is clear the alphabetic list is important as a source of possible entry words that could be selected for inclusion in a bilingual dictionary. In other words, alphabetic lists generated from the parallel corpus can assist the lexicographers in the process of lemmatisation. In lemmatised lists, only the basic form will be entered as a headword, for example, ‘account’. The headwords in most linguistic dictionaries are arranged in alphabetic order, so that this data would be helpful for an English-isiXhosa dictionary of finance terms. In analysing these headwords, they were verified in *GDX*, Volume 1 (A-J) 2006 and it was found that *iakhawunti* is lemmatised under letter ‘A’ but presented as the full form in the alphabetic list:

Example 1: *i·akhawunti*



i·**àkháwúnti** b/n 9/10 < Eng account:

1 ucwangciso-zimali ezingenayo nezichithwayo encwadini:
2 ityala, isikweliti:
3 isiphetshana sesikweliti:
4 imali eselugcinweni ngegama lomntu kuvimba, ebhankini:

1 debit and credit account, eg in a ledger;
2 bill, account, debt;
3 account, bill, debt;
4 a credit account in a bank that is in one's name.

1 rekening, bv in 'n grootboek;
2 rekening, staat van skulde;
3 rekening;
4 'n kredietrekening van 'n bank wat is in jou eie naam.

Page 1 of 1 0 words English (South Africa) 140

Figure 4.2: Headword iakhawunti from GDX

The above example illustrates the lemmatisation strategy that is commonly used in isiXhosa, where the stem tradition is commonly preferred. Svensén (1993:66) describes a headword as the heading of a dictionary entry. This is the first access structure which the users look up in order to find the various types of information in a dictionary. The alphabetic lists as an inventory list could aid the lexicographer in selecting a suitable approach for alphabetisation: letter by letter or word by word. In order to discover equivalent translations in the parallel corpus, the search feature of ParaConc is used as shown in the next section.

4.3 Searching for equivalent translations

As observed previously, a bilingual dictionary represents two languages. In a parallel corpus, the process is swift. A search word is entered and all possible translations are provided as illustrated in the figure below.

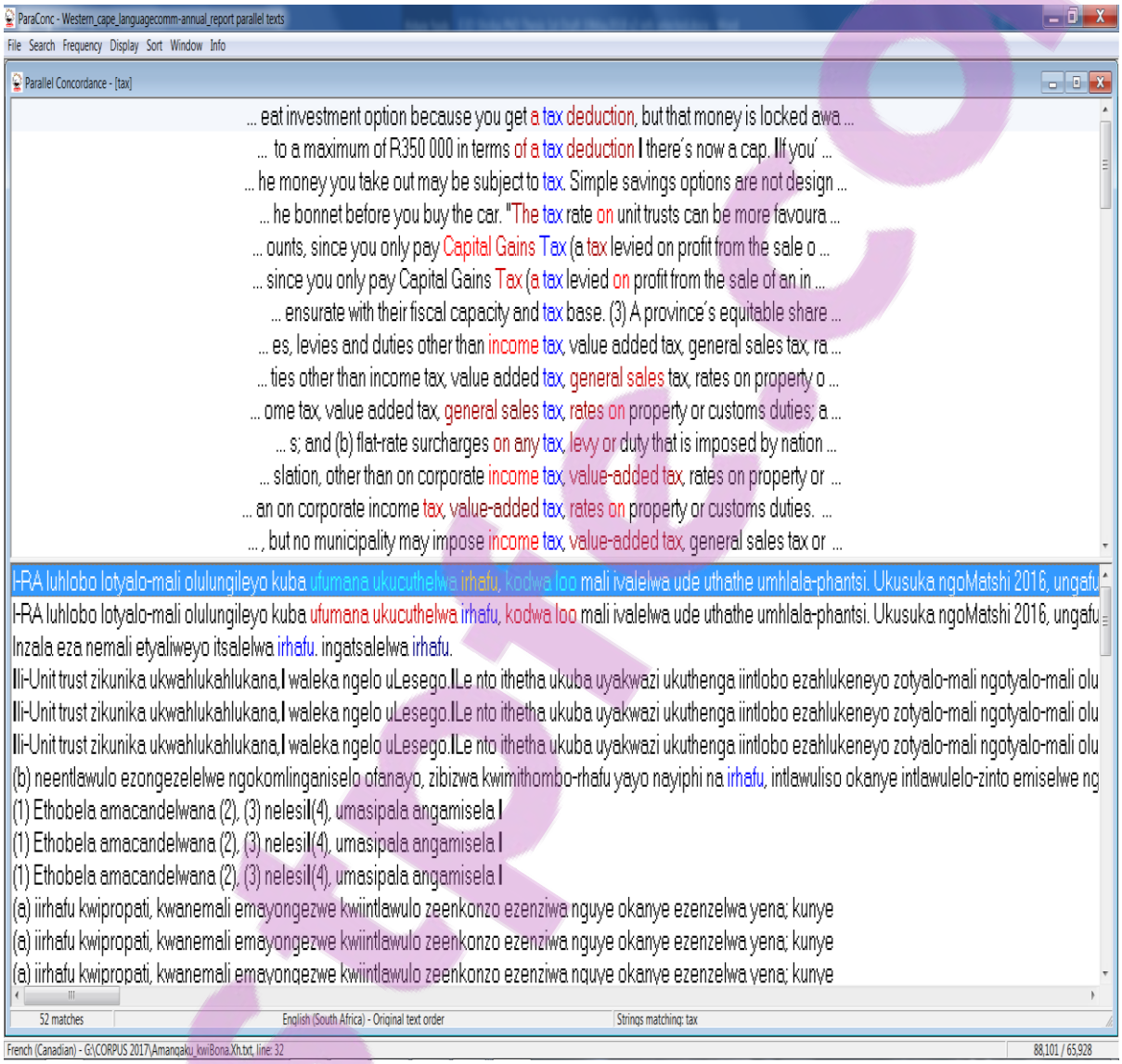


Figure 4.3 Display of translations of tax

Figure 4.3 displays how the search word 'tax' is uniformly translated as *irhafu* in the parallel texts. All the instances of the node are clearly marked in the upper window with their translations in the source language. These results indicate that 'tax' has the standard equivalent in the target language. This consensus is very significant as it will assist the lexicographer in selecting the possible translation equivalent in the absence

of any variants in the corpus evidence. The next section discusses how translation equivalents were extracted from a parallel corpus using ParaConc.

4.3.1 Identification of translation equivalents

The translated texts that constitute the EngXhPC are meant for communication between the writers and the target readers. The translation equivalents selected by the various translators are intended to meet the receptor's needs. Although literature in Chapter 2 revealed different views on the concept of translation, the general consensus is that translation is an interlingual process that aims to provide the equivalents of the source language in the target language. The literature reviewed indicated a heated controversy on the notion of equivalence in both translation studies and lexicography. The function-based theories both in lexicography and translation studies demonstrated a shift in how the notion of equivalence can be understood and applied in a communicative environment. The aim of the parallel texts that constitute the corpus under interrogation was communication of information in the domain of finance; therefore, one to one correspondence between the translation units and the texts was fundamental.

The primary task of a bilingual dictionary is to provide translation equivalents. In most cases the dictionary users consult this type of dictionary in order to access these. Translation equivalents according to Gouws and Prinsloo (2005) are target language lexical items that may be used to substitute the source language item in a specific situation. The corpus evidence from the EngXhPC provides an alternative approach for retrieving appropriate translation equivalents from the parallel corpus in context and in a specific situation. This relieves the lexicographer from the trouble of deciding on the equivalents because the translators have done that already. As noted, the use of translated parallel texts makes the whole exercise faster and more accurate. In Chapter 3, Table 3.1 examples of translation equivalents extracted from the EngXhPC reveal different degrees of equivalence that would be relevant for a specialised bilingual dictionary. The three types of equivalent relations that are identified in the EngXhPC are full equivalents, partial equivalents and zero equivalents. These are discussed below as they form an important information category in a bilingual dictionary.

4.3.2 Full equivalents

The findings of the study show that equivalence is a contextual matter particularly in subject specific domains. In the analysis of one-to-one equivalent relations tabulated in Figure 4.2 above, both frequency lists for instance illustrate a translational equivalent that provides an item of the source language (SL) with the same value as the target language (TL). In simpler words, in full equivalence, 'the source language item and the target language item have exactly the same meaning, function on the same stylistic level and represent the same register' (Gouws and Prinsloo, 2005:154). The aim is to match the meanings of the two lexical items. For example, the term 'expenditure' in English is translated as *inkcitho* in isiXhosa; this results in full equivalence. In the *Oxford South African Concise Dictionary*, expenditure means the action of spending funds or the amount of money spent. In its equivalent translation in *Oxford English-Xhosa Dictionary,* expenditure is treated as the sub-lemma of expend (*ukuchitha*, to use up) and translated as *inkcitho, indleko.* The table below presents selected examples of full equivalence that were drawn from the English-isiXhosa Parallel Corpus.

Table 4.3: Examples of full equivalents extracted from EngXhPC

English source term	***IsiXhosa target term***
1. Investment	*Utyalo-mali*
2. Debt	*Ityala*
3. Tax	*Irhafu*
4. Payment	*Intlawulo*
5. Interest	*Inzala*
6. Money	*Imali*
7. Shares	*Izabelo*
8. Income	*Ingeniso*
9. Surplus	*Intsalela*
10. Expenditure	*Inkcitho*

The table above provides common finance terms in English with their translational equivalents that indicate full equivalence. The English source word ‘debt’ in financials simply means the money that someone owes. The equivalent that shows the exact translation is *ityala*. In analysing the corpus evidence, ***example 9*** ‘surplus’ in Table 4.5 above, was looked up in two different dictionaries. In the *Oxford English-Xhosa Dictionary*, the translation equivalent for ‘surplus’ is *intsalela* which is preceded by an explanation (an amount of money that remains after needs have been supplied). ‘Surplus’ in the *Oxford South African Concise Dictionary* is defined as ‘an amount left over when requirements have been met’. These reference works confirm that the translation equivalent has the same value and meaning as the item from the source text. The fact that the translators used it uniformly in the various texts further demonstrates that this is a natural equivalent in isiXhosa. The frequency of surplus in the English corpus is 85, whilst *intsalela* occurs 23 times in isiXhosa. This difference in frequencies is caused by the concords affixed to the search term which are identified differently by ParaConc. For example, the translations of amount when influenced by concordial agreement are: *ingxowa-mali* ***ye****ntsalela* (surplus fund) or *iintlawulo* ***ze****ntsalela* (surplus payments). The two affixes preceding the stem are possessive concords according to different noun classes. The concordance information indicates that the translators are in agreement in providing the equivalent *intsalela* in the target language.

In Table 4.3, the ‘expenditure’ that appears 262 times in the English corpus is translated as *inkcitho* in isiXhosa. The latter term appears 186 times in the Xhosa corpus which shows that it is frequently used by most translators. This is another good example of what is seen as full equivalents identified from the parallel corpus where the source term means exactly what is given in the target language. The discrepancy in the frequency counts between the SL and TL word is due to the fact that isiXhosa is written conjuctively, as indicated in the literature review chapter. In context *inkcitho* may be found in the company of other words like *yenkcitho*. In the EngXhPC, the low frequency in isiXhosa is a notable trend. The following examples illustrate how example 10 ‘expenditure’ and its translation *inkcitho* are defined in different types of dictionaries that were selected by the researcher:

a) **expenditure 1.** the action of spending funds ►the amount of money spent. **2** the use of energy or other resources. (*Oxford South African Concise Dictionary* 2013:410).

b) **Expenditure** The costs or expenditure incurred by an organization. They may be expenditure or revenue expenditure… (Law, 2010:180).

c) *i·**nkcitho*** b/n 9/10:
1 *imali, impahla ekhutshelwe, ehlawulelwe, echithelwe into ethile*:
Back translation: Money, goods that have been paid or spent for something.
2 *ukusetyenziswa kwento enexabiso ngendlela engeyiyo, imosharha*: (*GDX Volume 2 (K-P*) 2003:627).
Back translation: To use something valuable in an irresponsible manner, waste.

The source language definitions of expenditure are taken from a general-purpose dictionary (a) and specialised dictionary (b), whilst the target language definition is extracted from the isiXhosa column of *GDX*. Sense 1 of the three references exemplifies full equivalence and the different translations from the EngXhPC indicate the absolute correspondence between the source language and the target language lexical items. The translators selected this type of equivalence because the word exists in the TL lexicon and is recorded in the dictionaries as illustrated in the examples above. This is to ensure that the translation is not only natural and fluent but also helps the receptor to understand the specialised source text. The options evidenced by corpus data encourages lexicographers to consult primary and secondary sources when deciding what to include in a dictionary.

The above examples indicate that complete/full equivalents can be identified from the EngXhPC; as such, parallel corpora can contribute positively to the development of specialised bilingual dictionaries. However, it was noted that the frequency of the lexical items that fulfil full equivalence is low in the EngXhPC because one-to-one correlations are difficult to find between developed and underdeveloped languages. Nevertheless, it is important to state that full equivalence contributes positively to the standardisation of terminology in isiXhosa.

Additionally, the identified equivalents could make a valuable contribution in the bilingual dictionaries because they promote naturalness and fluency in the translated

texts. Their inclusion in a LSP dictionary could improve the communicative role of such a dictionary. The selection of complete equivalents from this parallel corpus will result in the production of user-friendly bilingual specialised dictionaries where target users will be accessing information that is understood in their target culture. The users will not only understand the concepts but will also be in a position to reproduce them. But, as stated above, full equivalence is not always possible between developed and underdeveloped languages, leading to other types of equivalents such as partial equivalents which are discussed below.

4.3.3 Partial equivalents

In an attempt to produce target-oriented texts, translators always look for appropriate equivalents when translating. Dictionary users have the same need: whenever they access a bilingual dictionary they expect to retrieve suitable equivalents for the various problems they encounter. The researcher found that in the English-isiXhosa corpus, when full equivalents could not be found, some translators opted for partial equivalents. Partial equivalence according to Gouws and Prinsloo (2005:155) 'prevails where the source and target language items do not display a one to one correspondence'. In order to fill this gap an equivalent with the same function as the source language may be used. The meaning of the translation equivalent corresponds partly to its counterpart in the source language. The table below shows some of the examples drawn from the EngXhPC.

Table 4.4: Selected examples of partial equivalents in context

Source text	**Target text**
1. finance	*Imali*
2. cash	*Imali eziinkozo*
3. liabilities	*Amatyala*
4. dividend	*Inzala*
5. Treasury	*uNondyebo*

The two terms ‘finance’ and ‘cash’ demonstrate good examples of partial equivalence. The following examples are extracted from the concordance view of the EngXhPC:

Example 1:

> **English ST**: Your spending habits play a big role in shaping your **finances.**
> **IsiXhosa TT:** *Indlela osebenzisa ngayo **imali** idlala indima enkulu ekubumbeni imeko yakho **yezemali***.

The term ‘finance*’* does not reflect one to one correspondence with the equivalent *imali.* It conveys a broader meaning than the term money; however, corpus evidence confirms that finance is translated as *imali* in isiXhosa and is usually used in plural form. In a bilingual dictionary of English and isiXhosa money is translated as *imali* whereas finance has a target equivalent of *imali: indyebo*. *Imali* when referred to as ‘finance’ displays partial correspondence. A monolingual dictionary provides a clearer distinction between the two terms. According to the *Cambridge International Dictionary of English* (Procter, 1995:520) finance is a supply of money and finances, in other words, denotes money available to spend. The same dictionary defines money as ‘the coins or notes which are used to buy things or the amount of these that one person has’. The *South African Concise Dictionary* (2010:435) on the other hand provides another broader definition of finance as ‘the management of large amounts of money, especially by government or large companies or monetary support for an enterprise’ and finances as ‘monetary resources and affairs of the state, organisation, or person’. In the definitions ‘finance’ refers to the manner in which money is controlled and money is a form of financing.

In example 1, “finances’ for instance, is used broadly to mean managing all one’s finance issues which may include being able to budget, save, spend etcetera. The second example shows that cash is also translated as imali:

Example 2:

> **ST:** “Put extra **cash** into your savings account or your bond…’’
> **TT:** “*Faka **imali** egqithisileyo kwi-akhawunti yakho yokonga okanye kwintlawulo yendlu*…’

‘Cash’ in example 2 depicts a partial relation when translated as *imali* in isiXhosa. Cash is money in the form of coins or notes. Some translators tried to avoid choosing this equivalent and opted for *imali eziinkozo okanye esesandleni* (Money in coins or at hand). The following sentence illustrates the use of an appropriate equivalent that represents an exact equivalent of the source language word ‘cash’:

> ST: **Cash** and cash equivalent at the beginning of the year.
> TT: ***Imali eziinkozo*** *nokufana nayo ekuqaleni konyaka*.

The decision to opt for a partial equivalent in example 1 may be attributed to lack of an exact equivalent or a hyponym, so that translators selected an equivalent that has a proportional meaning to the source term, whilst in example 2, the translator selected a term that is lexicalised in isiXhosa to avoid the loan word *ikheshi.* In example 5, in Table 4.6 the word treasury with capital ‘T’ refers to a government department whose responsibility is to oversee finance and policies on financial matters. The equivalent given in isiXhosa is *uNondyebo*. The translators of the parallel texts use either *uNondyebo* or *uNongxowa-mali*. In isiXhosa *unondyebo* or *unongxowa* is a person who is responsible for keeping the finances and assets of the organisation; consequently, these target equivalents do not really mean that Treasury is keeping the monies. Nonetheless the functions of the treasurer and Treasury are similar. It is therefore evident that instances of this kind of equivalence are easily depicted when a basic search tool of ParaConc is performed. Scrolling down the concordance lines in both windows provides the different equivalents of the search term and their statistical information. The context in which these words are used will help the lexicographer to make well-informed decisions in the choice of appropriate equivalents. The following sub-section will consider the situation where neither full nor partial equivalents were selected by translators, and thereafter analyse these kinds of equivalents in the EngXhPC through ParaConc.

4.3.4 Zero/Non-equivalents

The translators of the financial texts did not select one type of equivalent during translation. Because of the lexical gap between English and isiXhosa, the lack of equivalents or no equivalence is obvious in the finance domain. When translators could not find a full or partial equivalent, they opted for either a loan word or an explanatory equivalent so that a clear message is conveyed to the readers. Gouws and Prinsloo (2010) and Gouws (2002) call this surrogate equivalence, an equivalent relation that results in different types of surrogate equivalents. These equivalents are represented by either the use of pure loan words or indigenised loan words; these are discussed below. These terms are prevalent in EngXhPC.

4.3.4.1 Indigenised loan words

These loan words are a result of borrowing, a translation strategy that is often devised by the translators to solve the non-equivalence that exists between English and isiXhosa. Borrowing, according to Alberts (2013), is the process whereby new words are formed by adopting words from other languages. Borrowing, according to Satyo as cited by Moropa (2005), has played an important role in the development of isiXhosa. Loan words in the EngXhPC can be displayed in the frequency lists and concordance lines. More examples of loan words extracted by ParaConc from the parallel corpus are listed in a table below.

Table 4.5: Examples of indigenised loan words extracted from the parallel corpus

Word count	English word	Word count	isiXhosa translation
426	Assets	38/51	*iiasethi/ii-asethi*
321	Statements	18	*Izitetimenti*
233	Cash	24	*Ikheshi*
188	Capital	1	*Ikhapitali*
99	Insurance	3/21	*i-ishorensi/inshurensi*
95	Benefits	19	*Ibhenefithi*
83	Budget	5	*Ibhajethi*
70	Property	23	*Ipropati*
66	Company	29	*Inkampani*
61	Accounts	3/5	*iiakhawunti/ii-akhawunti*
53	Vote	78	*Ivoti*
52	Balance	3	*Ibhalansi*
48	Market	6/1	*imarike/imakethi*
30	Rand	2	*Irandi*
30	Virement	12	*Ivayamenti*

The examples in Table 4.5 above illustrate indigenised loan words. Although these words are borrowed from English, the orthography suits the writing and spelling rules of isiXhosa. Indigenised loan words are usually entered in a dictionary only if there are no other forms of equivalents available: for example, the term budget has a translational equivalent *uhlahlo lwabiwo-mali*. The translator, however, preferred *ibhajethi* instead of *uhlahlo lwabiwo-mali* perhaps because it is the most frequently used term in the spoken language. In a dictionary entry, for instance, *ibhajethi* can be listed as an alternative form after the target language item. Mojela (2010) echoes that

loan words can be used as synonyms for the coined indigenous lexical items. In the case of 'bank'/*ibhanki* the use of a loan word is justifiable because there is no TL equivalent for the concept. Indigenised loan words are adapted to the writing system and some of them cannot be recognised as loan words anymore because of consistent use. The ordering of these words also depends on the function and prospective users of the dictionary. The following section discusses pure loan words in dictionary making.

4.3.4.2 Pure loan words

A pure loan word is a word that is taken over from the source language to the target language without changing its phonological and morphological structure (Mtintsilana and Morris 1988). Words like these are also referred to as direct loanwords because they are borrowed directly from the source language with little or no modification. According to Moropa (2005) the pure loan words in isiXhosa retain their original structure except for the prefixes which are governed by the concordial agreements of isiXhosa as a conjuctive language. During this adaptation, the meaning of pure loan words is also not disturbed, but remains intact. The translators normally use this strategy when the source term lacks an equivalent in the target language. It is employed as a last resort compared to indigenisation. Table 4.7 illustrates 8 examples of pure loan words that were retrieved from the parallel corpus. Through the search function, the following loan words were identified.

Table 4.6: Examples of pure loan words in context

1. Use your **credit card** wisely.	*Sebenzisa **icredit card** yakho ngobulumko.*
2. Ask the **medical aid** the things that are paid for.	*Buza **imedical aid** yakho ukuba zintoni ekubhatalela zona.*
3. You cannot buy a collective investment scheme in property with a monthly **debit order**	*Ngeke uthenge isikim sotyala-mali seqela nge-**debit order**.*
4. If your goals are short-term, then something like a **MoneyMarket** account offered by almost every bank, is a good option.	*Ukuba iinjongo zakho zezexesha elifutshane, ngoko ke i-akhawunti **yeMoneyMarket** enikezwa phantse zizo zonke iibhanki, luhlobo olulungileyo.*
5. Buying and selling the participatory interests is like buying and selling in a company listed on the **JSE Limited.**	*Ukuthenga nokuthengisa le nzala yotyalo-mali lweqela ifana nokuthenga izabelo kwinkampani ekuluhlu lwe-**JSE Limited**.*
6. The entity has reported monthly and quarterly to the Provincial Treasury as is required by the **PFMA.**	*IQumrhu linike ingxelo ngenyanga nangekota kuNondyebo wePhondo ngokweemfuno ze-**PFMA.***
7. The **Total Expense Ratio** (TER: the total cost of a fund to the investor) must be considered.	***I-Total Expense Ratio** (TER: lixabiso lemali kumtyali-zimali) kufuneka ifakwe ezingqondweni.*
8. Standards of **GRAP** approved but not yet effective for the reporting period:	*Imigangatho **yeGRAP** evunyiweyo kodwa ingekasetyenziswa kulo nyaka kwenziwa ngawo ingxelo*:

The information in Table 4.6 above consists of a few examples of pure loan words that were extracted from the EngXhPC. The search facility of ParaConc is used to display these loan words in the KWIC format. The loan words represent different parts of speech such as compounds (see Examples 1, 2, 3, 4, 7), proper nouns (example 5), acronyms (example 8), abbreviations (5, 6) and the like. Although these words have retained the original form, the prefixes or concord are attached before the English word in line with the morphosyntactic rules of isiXhosa as discussed in Section 2.8 of

Chapter 2. According to the noun class system, loan words are classified under class 9/10 (singular) and the plural take the class prefixes i-/ii as shown in example 1 and 2 above. There is no hyphen between the class prefix and the borrowed word. The translator has followed the orthographic rules of the target language (see also Table 2.1 in Chapter 2) in presenting the pure loan words.

In the first example, the term 'credit card' is retained in its original form in the target language isiXhosa, but preceded by the class prefix i-. A credit card is 'a plastic card issued by a bank or finance organization to enable holders to obtain credit in shops, hotels, restaurant, petrol stations, etc.' *(Oxford Dictionary of Finance and Banking,* 2014:112). This results in one-to-one equivalence: however, although pure loan words are suitable for disseminating information, to include them like this in a bilingual dictionary would not benefit a dictionary user who is looking for semantic information in the target language. In the parallel texts that are meant for communicating specialist information, such equivalents are appropriate because the meaning is not diluted nor confusing. The inclusion of pure loan words as translation equivalents needs to be carefully considered, looking at what the potential users might need from the dictionary.

Although borrowing is regarded as a term formation strategy to develop African languages, scholars have some reservations when it is used despite the existence of target terms. Gumbo (2016) is of the view that some of the words are cumbersome, artificial and vague. In example 7 above, after the pure loanword, *iTotal Expense Ratio* the information enclosed in brackets explains that TER is the amount of money to the investors. Without the explanatory note, 'TER' as the dictionary entry would be vague for a target user who does not have financial background. The researcher verified the word from three dictionaries, namely, *The Penguin International Dictionary of Finance* (Bannock and Manser, 2003), *Oxford Dictionary of Finance and Banking* (Law, 2014) and *Collins Dictionary of Business* (Pass, Lowes, Pendleton and Chadwick,1995). The dictionary search came with zero results. According to Wikipedia, 'the total expense ratio, or TER, is a measure of the total cost of a fund to the investor'; https://en.wikipedia.org/wiki (accessed on 1 May 2018). The non-inclusion of this word in the dictionaries consulted indicates that it is a new word in the finance, banking and business sector. By the explanation, the translator was trying to simplify the new word so that the reader could benefit from a clearer understanding. It is suggested in this study that pure loan words that may cause problems in the comprehension of financial

information should be included with an explanation. Another form of equivalence that was noted in the study was the use of explanatory equivalents.

4.3.4.3 Explanatory equivalents

To address non-equivalence an explanation or paraphrase could be used as a form of explanatory equivalent. A paraphrase is a definition or explanation provided to overcome non-equivalence. Ndhlovu (2012:147) explains that 'paraphrasing is used when the target language lacks a concept presented by the source term'. In the EngXhPC, explanatory equivalents are commonly used by translators to resolve the problem of non-equivalence (see Table 4.7 below).

Table 4.7: Examples of explanatory equivalents

Source Text	**Target Text**	**Back translation**
1. Accruals	*iimali ezongezelelweyo*	*Monies added.*
2. Annuity	*inxenye yomhlalaphantsi*	*Portion of pension.*
3. Cash	*imali eziinkozo*	*Money that is coins.*
4. Consumers	*abasebenzisi beenkonzo*	*Users of services.*
5. Executor/executrix	*umabi welifa, umabi wempahla*	*The one that divides an estate.*
6. Creditors	*abantu ababanjwa imali*	*Those owned money.*
7. Fiduciary	*Umntu ogcina ipropati*	*One who keeps property.*
8. Investing	*ukutyala imali*	*To sow money.*
9. Testator	*umntu owenze umyolelo*	*One who has drawn up a will.*
10. Receivables	*Imali elindelweyo*	*Expected money.*

As observed in Table 4.7 above, the selected examples illustrate the employment of paraphrasing as a term formation process when translators have failed to identify an exact target equivalent. Paraphrasing involves the use of descriptions or explanations.

These explanations or paraphrases give the users a clearer understanding of the meaning. Scholars in translation studies have noted the advantages and disadvantages of paraphrasing (Sineke, 2005). Madiba (2000:220) views it as a productive way of engineering knowledge and of easily describing the meaning of foreign words. The present researcher supports the inclusion of explanatory equivalents in bilingual specialised dictionaries in isiXhosa as one of South Africa's indigenous languages. The inclusion of paraphrases of meaning is very effective in bilingual dictionaries because they provide definitions that enhance the semantic information and the growth of isiXhosa as a scientific language. Explanatory equivalents are very helpful in avoiding ambiguity, when a foreign word is used, though they are sometimes included in closed brackets for guiding the user in accessing the appropriated information.

In a nutshell, translation equivalents extracted from the parallel corpus clearly reveal that exact equivalence between English and isiXhosa finance terms is not always possible. The corpus data reveals different equivalence relations selected by translators in order to achieve translational equivalents. The lack of equivalence is caused by the lack of standardised financial terms and resources such as dictionaries that are specialised and current. Meaning is what is looked for by the dictionary users and the prime purpose of a bilingual dictionary is to provide different types of translation equivalents in order to facilitate optimum communication.

Using ParaConc in interrogating the EngXhPC confirmed the words of Teurbet (2002:204) that parallel corpora are 'repositories of translation units and their equivalents in the target language which can be applied for selection and verification of translation equivalents in a bilingual dictionary'. The pool of equivalents extracted by the concordance will facilitate the process of selecting them. The method is quicker and more efficient as terms and their translations are provided within seconds. The translators have carried out their research to find authentic and usable equivalents, while the lexicographer has to focus on writing the dictionary entries based on the available data. It is however important to note that no corpus is foolproof, which means that lexicographers may have to go beyond the corpus to authenticate facts. This includes consulting field experts who work with these terms on a daily basis. Previous dictionaries are also important resources when creating specialised dictionaries as some specialised terms may be found in these dictionaries.

Another feature of ParaConc that is important in dictionary making is the frequency list. This feature is necessary in indicating the frequencies of words in a particular corpus, as illustrated below.

4.4 Frequency list

Frequency lists are lists that are generated by a corpus analysis tool according to the number of occurrences of terms in the various texts of the corpus. That is, they show how many times a word appears in a particular corpus, be this the English corpus or isiXhosa corpus. A number of researchers have alluded to the role of frequency lists in corpus-based translation studies and lexicography (cf. Bennet, 2010; Ndhlovu, 2016, amongst others). In corpus-based lexicography, frequency counts can assist in developing lemma lists of appropriate words. They aid a lexicographer in deciding what to include in or exclude from a dictionary. Such counts may also help in the revision of existing dictionaries by selecting new words that are crucial in updating the dictionary. The following screenshot shows the frequency lists generated from the EngXhPC. On the left of the screenshot is a list of English words that exist in all the 10 corpus files uploaded, while on the far right the screenshot displays a frequency in isiXhosa (French Canadian).

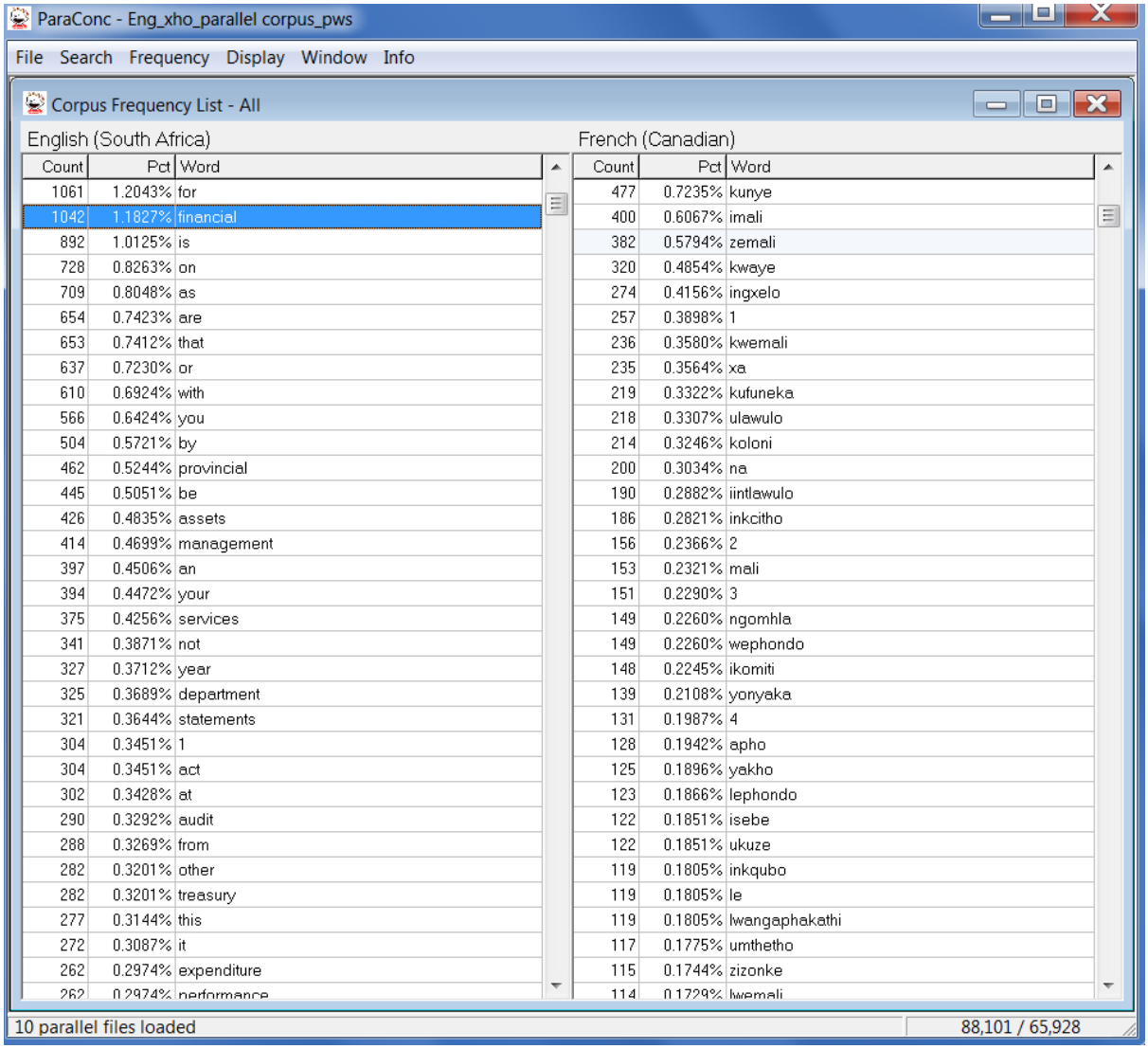


Figure 4.4: Frequency list from English-isiXhosa parallel corpus

In the parallel corpus above, the most frequent words in English and isiXhosa happen to be grammatical words. In the English frequency list, the preposition ‘for’ comes up high on the list, whilst in isiXhosa the most frequent word is the conjuctive *kunye*. The two words do not hold meaning when alone, because their role is to connect other words in a sentence, hence they are called function words. In the examination of context in the concordance they play a significant role as they form part of collocations and also constitute the lexical structure of the EngXhPC.

In answering the research question, the analysis will focus on content words that are finance based. The lexical words as seen above are words that have a meaning. The adjective ‘financial’ is the most frequent finance word in the English list with a frequency

count of 1042, whilst in isiXhosa, the noun *imali* with the source term ‘money/fund/finance/cash’ comes up with a frequency of 400. The possible equivalents of financial in the target list are *zemali* with 382 frequency counts and *kwemali* with 236 counts. This output suggests that these two words are common words in the texts and in the financial domain; therefore, a lexicographer will not be able to leave them out. The second term with high frequency counts of 426 in the parallel corpus is the financial term ‘assets’. In the *Oxford English-Xhosa Dictionary*, it is lemmatised in basic form but additional information that guides a dictionary user is included:

> **asset**, n. (usu.pl) *izinto umntu anazo ezinokuthengiselwa ukuhlawulwa kwamatyala akhe;* (valuable or useful quality or skill)

In the corpus frequency list, the financial term ‘assets’ is more frequent than the basic form, and in the above dictionary entry this is confirmed by the information in brackets which explains that it is usually used in the plural form: the frequency count from the EngXhPC attests to this, indicating how reliable and authentic corpus data is.

The above examples indicate that it is possible to obtain an objective selection of headwords or lemma by means of ParaConc and are useful in building the macrostructure of both languages, that is, English and isiXhosa. The macrostructure is the ‘overall list structure which allows the compiler and the user to locate information in a reference work’ (Hartmann and James 2001:91). The frequency list extracted from the EngXhPC confirms that the parallel corpus consists of finance terms. This is in line with what Li (2006 in McEnery and Xiao 2016) says: that dictionaries which are corpus-based could ensure a systematic coverage of headwords of practical value, accurate definitions and appropriate illustrative examples. The lists provide relevant information for a specialised dictionary. The provision of such headwords will be beneficial for a variety of potential users, such as specialists, general public, translators, researchers, students and others. This will not only help in the identification of target equivalents but is able to provide an ordered list in isiXhosa. The table below provides a summary of the first 5 frequent financial terms for each language. The high frequency counts qualify the terms to be included in the lemma list; these stand a high chance of being included in the dictionary.

Table 4.8: Top 5 frequent words extracted from EngXhPC

Count	Percentage	Word	Count	percentage	Word
1042	7.3609%	Financial	**400**	0.6067%	*Imali*
426	0.4835%	Assets	**382**	0.5735%	*Zemali*
290	0.3292%	Audit	**236**	0.3580%	*Kwemali*
282	0.3087%	Treasury	190	0.2882%	*Iintlawulo*
262	0.2974	Expenditure	186	0.2821%	*Inkcitho*

A lemmatised frequency list assists in ensuring that the headwords that are normally looked up by the dictionary users are included in the dictionary. The advantage of frequency according to Gouws and Prinsloo (2005:30) is that it ensures that frequently used words are not accidentally omitted and that precious dictionary space will not be taken up by the lemmas less likely to be consulted by the target user.

The frequency lists in a parallel corpus could also be of help in capturing new words that are introduced in a specific language. That is, frequency lists are useful in the design of an inclusion policy for a particular dictionary. In modern times, frequency-based dictionaries are mostly preferred because they only focus on needed information. In the conventional methods of dictionary making it is a common error to omit very frequent words. Frequency counts are products of corpus analysis and are not possible manually. In this study, ParaConc was used to count all the occurrences of words in all the corpus texts (see Appendix 4- frequency lists from EngXhPC).

4.5 Generating keyword/hot word lists

The process of extracting hot words is explained in Section 3.5.7 of Chapter 3. As stated previously, hot words are defined as the words which occur with an unusually high frequency in a text or corpus (Bowker and Pearson, 2002:114). Hot words are also referred to as keywords. The development of a keyword list can provide another kind of valuable analysis. The hot word tool in a parallel corpus identifies all the words that indicate an unusual frequency in relation to the source word. In fact, it shows the very important words that distinguish one corpus or one text from another. The keywords are suggested by the corpus analysis tool. The following is a list of keywords generated from a concordance view of the word investment.

Table 4.9 Hotwords for the word 'investment'

Rank	**Word**
81.19	*utyalo-mali*
74.41	*sotyalo-mali*
61.02	*zotyalo-mali*
52.16	*Seqela*
49.13	*Lweqela*
43.13	*Kwipropati*
33.10	*lotyalo-mali*
27.61	*yotyalo-mali*

The most highly ranked word in the above keyword list is *utyalo-mali*, an equivalent translation of the source word 'investment'. The keywords in the list are arranged according to their hotness or keyness (cf. Barlow, 2003; Moropa, 2005; Ndhlovu, 2012). Beside the translations, one can also identify the collocations of translations of the search word and the morphological structure of the words that are related to the search word. For example, the words that occur with investment/*utyalo-mali* are: *seqela*, *lweqela*, *kwipropati.* The list indicates the key words that are most significant in the parallel text or corpus. The information on the morphological patterns of the stem

–*tyalo-mali* is also picked up in the keyword list as in ***yo****tyalo-mali*, ***so****tyalo-mali* etc. The bolded formatives are the possessive concords and the translations of the source word 'of', a preposition. The source phrase 'of investment' is translated as *yo-/sotyalomali* to indicate a concordial agreement. The function word 'of' is written disjunctively from the noun 'investment' because English is not agglutinative. The linguistic nature of the two languages is easily uncovered**.** Barlow (2003) sums up the significance of this analysis as a simple attempt to deal with morphological variation without resorting to language specific resources. The analysis of keywords and frequency lists helps in determining the content and the lexical structure of the EngXhPC. The comparison of the frequency lists and keywords would assist in the selection criteria during dictionary writing. The lexicographer could use the information to identify the lemma candidates with their frequency information (see also *Appendix 4* for a frequency list developed from this parallel corpus).

4.6 Identification of synonyms

Synonyms constitute an important form of data in bilingual dictionaries: 'A synonym is a word that has the same meaning (in a particular context) as another word' (Crystal, 2010:109). The use of the EngXhPC made it easy to identify synonyms through the search facility and hot words list. The following table is an illustration of common synonyms that are extracted from the EngXhPC by means of ParaConc:

Table 4.10: List of common synonyms in the parallel corpus

Source term	**Target term**
Salary	*umvuzo, umrholo*
Pay	*ukuhlawula, ukubhatala*
Save	*ukonga, ukulondoloza*
Treasurer	*unondyebo, unongxowa*
Budget	*uhlahlo lwabiwo-mali, ulwabiwo-mali, ibhajethi*

The contexts of the target language aid the analyst to identify synonymy between the source language items and its equivalents in the target language. These synonyms were identified by searching the SL term; the concordance lines displayed how it is translated in the parallel corpus. The following examples resulted from the term search.

a) **Salary > *umvuzo/umrholo***

ST: For most women the UIF maternity benefit does not match up to their **salary**.

TT: *Kumabhinqa amaninzi imali ye-UIF ayilingani **nomrholo** wawo.*

ST: You and your employer contribute a certain percentage of your **salary** monthly towards this retirement fund.

TT: *Wena nomqeshi wakho nirhuma ipesenti ethile **yomvuzo** nyanga nganye iye kwingxowa-mali yomhlala-phantsi*.

b) **pay > *ukuhlawula/ukubhatala***

ST: It might be a good idea to **pay** school fees upfront when you have money.

TT: *Yingcamango entle **ukubhatala** iindleko zesikolo ngoku usenemali.*

ST: Do I need to **pay** tax on profit?

TT: *Ingaba kufanele **ndihlawule** irhafu kwinzuzo*?

c) **save > *ukonga, ukulondoloza***

ST: The difference between savings and investments. Savings enable you to **save** your future and that of your family.

TT: *Umahluko phakathi kolondolozo notyalo-mali. Ulondolozo lukwenza **ulondolozele** ikamva lakho kunye nolosapho lwakho*

ST: Can you **save** on food, water, entertainment or other areas in your budget?

TT: ***Ungonga** ekutyeni, kwezolonwabo okanye kwenye into*?

d) **treasury > *unondyebo/unongxowa***

ST: National **Treasury** subsidised the audit fees of the Western Cape Language Committee

TT: ***UNondyebo** weSizwe uncedise ekukhupheni iintlawulo zophicotho-ziincwadi kwiKomiti yeeLwimi eNtshona Koloni*.

ST: National **Treasury** provides the guide to complete Annual Report.

TT: ***UNongxowa-mali*** *weSizwe unika izikhokelo zokuqulunqwa kwengxelo zoNyaka-mali*.

e) budget >*uhlahlo-lwabiwo-mali, ulwabiwo-mali, ibhajethi*

ST: Jet-setting on a budget

TT: *Yenza* ***ibhajethi*** *xa ukhenketha*.

ST: The approved **budget** covers the period from 1 April 2013 to 31 March 2014.

TT: ***Ulwabiwo-mali*** *oluvunyiweyo luquka ixesha elisuka kumhla woku-1 kuTshazimpunzi ngowama-2013 ukuya kowama-31 kweyoKwindla ngowama 2014*.

ST: Reconciliation between budget and cash flow statement.

TT: *Ungqamaniso phakathi kohlahlo-lwabiwo-mali nenkcazo yokuhamaba kwemali*.

The examples above show words that are different in form but have the same meaning. All these examples display exactness of meaning. The words that possess this quality are known as complete or full synonyms. They point to the richness of isiXhosa. For example, in sentence (b) 'pay' has the equivalent *ukuhlawula* and *ukubhatala. Ukubhatala* is a loan word from the Afrikaans word 'betaal', whilst *ukuhlawula* is a TL lexical item. In the context above the two target words mean exactly the same thing, to pay. The verb 'pay' in the *Oxford English-Xhosa Dictionary* means *–hlawula, -bhatala*. This dictionary entry confirms that the two target equivalents share the same qualities. The corpus evidence also identifies instances where language synonyms from more than one source are translated by one target language item. These are discussed below.

(a) expenses/costs ►*iindleko*

The two words are picked up from the parallel corpus as synonyms and, there, are used interchangeably by the translators. According to the *Oxford South African Dictionary* (2013:410) an "**expense** is cost incurred or required►**(expenses**) costs incurred in the performance of a job or task'.

Iindleko in *GDX Volume 2 (K-P)* (2003:478) in sense 1 is defined as *'inkcitho-mali ngentengo, ngokuhlawulwa kwamatyala nasezintweni ezifunekayo'*.

In both the English and isiXhosa definitions, expenses are costs, which clearly shows synonymy. Both terms: **cost** and **expenses** denote an expenditure. The evidence from this corpus confirms what is defined in the two dictionaries above.

Table 4.11: Revenue/income/profit > *ingeniso*

Source text	**Target text**
Equitable shares and **revenue** shares.	*Izabelo ezingeyonkqatho nezahlulo* ***zengeniso*****.**
A collective investment scheme in property brings current future **income** to the investors as follows:	*Isikim sotyalo-mali seqela kwipropati sizisa* ***ingeniso*** *yangoku neyexa elizayo kumtyali-mali ngezi ndlela zilandelayo*:..
Investor: Any person who purchases investment products with the expectation of achieving **profit** and/or income.	*Umtyali-mali: Nabani na othatha imveliso yotyalo-mali elindele ukufumana inzuzo kunye/okanye* ***ingeniso.***

Table 4.11, above, illustrates how English synonyms: 'revenue, income and profit' are translated with one equivalent *ingeniso* in isiXhosa. As indicated in Table 4.11 above, the contexts in which these synonyms are used demonstrate how translators of different texts have consistently selected *ingeniso* as the suitable equivalent for the three synonyms. Upon the analysis of the parallel corpus the word 'revenue' has 226 matches, *ingeniso* has 113 matches, whilst income records 62 matches. The word 'revenue'/*ingeniso* seems to be frequently used in the EngXhPC compared to income and profit. This information is very important for the lexicographer, because it can provide guidance on the inclusion strategy. The following table illustrates how SL synonyms, 'shares and dividends', are translated into the target equivalent *izabelo.*

Table 4.12: Shares/dividends > *izabelo*

Source text	**Target text**
How does one know which **shares** to buy? FSB brochures	*Umntu wazi njani ukuba ziziphi* ***izabelo*** *amakazithenge*?
Interest, **dividends** and rent on land. AFS Provincial TreasuryWC	*Inzala,* ***izabelo*** *nerente yomhlaba*.

The target equivalent, *izabelo* in Table 4.12 above, refers to both shares and dividends. A dividend is a payment by a company to its shareholders, usually as a distribution of profits (https://support.easyequities.co.za). According to Wikipedia, a share is an indivisible unit of capital, expressing the ownership between the company and the shareholder (https://en.wikipedia.org). The two definitions show similarities but also a slight difference between the two. This is characteristic of synonyms in general. In isiXhosa, their translation equivalent is *isabelo*. In the *Oxford English Xhosa Dictionary*, a dividend is explained as a 'payment of share or profit'. In the *Greater Dictionary of isiXhosa* (Volume 1 A-J: 2006:4), *isabelo* is defined as '*into owabelwe yona, isilunga, into elunge kuwe/nawe, isahlulo, isiqephu, umhlomlo, inxalenye, umphungulo, inxaxheba (*one's portion; one's share, portion of what is being apportioned'). In this definition, a 'share'*/isabelo is* synonymous with *'*dividend'*/isahlulo.* The use of the target equivalent *isabelo* in the EngXhPC for both words is confirmed by all the dictionaries cited above.

Table 4.13: Cost/value/price > *ixabiso*

Source text	**Target text**
The total **cost** of a fund to the investors must be taken into consideration. *Bona* Articles	***Ixabiso*** *lemali kumtyali-mali kufuneka lifakwe engqondweni*.
The **price** of an asset in the market.	***Ixabiso*** *le-asethi emarikeni.*
Market **value.**	***Ixabiso*** *lemarike*.

The table above illustrates the context and translations of three source words: cost, value, price which are consistently translated as *ixabiso* in the target language. *Ixabiso* generally refers to the amount charged, value or worth. This indicates that the translators have used other sources to confirm their translations. The following table presents how *umvuzo* is used in a context.

Table 4.14: Salary/remuneration/wages > *umvuzo*

Source text	**Target text**
The increase in **basic salary** is due to graduate interns that were appointed as contract workers and now earn a higher salary plus 37% for fringe benefits.	*Ukwanda **kumvuzo wesiseko** kubangelwa ngabamkela befunda abathe baqeshwa njengabasebenzi bethutyana yaye ngoku bamkela imivuzo ephakamileyo xa kudityaniswe nama-37% ezibonelelo.*
Salaries and **wages** are expensed in the statements of financial performance when the final authorisation for payments is effected in the system.	***Imivuzo** iyavela kwisitetimenti sokusebenza kwemali xa ugunyaziso lokugqibela lusenziwa kwinkqubo.*
Board member's ***remuneration*.** PanSalb AFS (2002)	***Imivuzo** yamalungu eBhodi.*

In the first sentence in Table 4.14, salary is translated as *umvuzo*, but due to the affix ku- the basic form has slightly changed. In the second context, wages are also translated as *umvuzo.* To avoid redundancy, the translator here has not translated salaries because in the target language it means the same thing. An example of four synonyms in English with one equivalent follows.

Table 4.15: Debt/credit/liability/bill > *ityala*

Source text	Target text
Debts are written off when identified as irrecoverable.	***Amatyala*** *ayacinywa xa egqalwe njengangenakuhlawuleka*.
Buying **on credit** also means that you could be living beyond your means.	*Ukuthenga* ***ngetyala*** *kuthetha ukuba usebenzisa imali ongenayo*.
Contingent **liabilities**.	***Amatyala*** *angacwangcwiswanga*.
Pay **funeral bills**, outstanding debts and valid claims against the estate.	*Ukuhlawula* ***amatyala*** *omngcwabo, amatyala ebengekahlawulwa anyanisekileyo kwimpahla yomntu*.

The table above indicates that *ityala* in isiXhosa is the equivalent of the multiple synonyms: debt, credit, liabilities and bills. In the following example one finds another interesting example of three other SL synonyms with one equivalent in the target language.

Table 4.16: profit/dividends > *inzuzo*

Source text	Target text
If a company makes a **profit**, more people will want to own shares in the company and the prices of that share will rise.	*Ukuba inkampani yenza* ***inzuzo,*** *abantu abaninzi baza kufuna ukuba ngabanini-zabelo kwinkampani yaye aya kunyuka amaxabiso ezabelo.*
Receive payments that are due to the estate such as interests, **dividends,** investments and other income…	*Ukufumana iintlawulo ezilindlelke kwizinto zomntu ezifana nenzala,* ***inzuzo,*** *utyalo-mali kunye nenye ingeniso…*

The synonyms behave differently in certain contextual environments. In Table 4.12 above, dividends are translated as *izabelo* and in the above table, interestingly, the word has taken an equivalent ***inzuzo.*** It is noticeable in all the contexts in which these synonyms occur that they do not mean the same thing every time; in other words, when lexicographers select synonyms for inclusion in the dictionary the headword and the context should be carefully examined. 'The use of common synonyms has been widely used by Xhosa translators when translating financial texts' (Moropa, 2005:134).

It is not always straightforward to find exact synonyms in the languages, since absolute synonyms are very rare. The use of common synonyms is intended to ensure that the translation is functional in order to satisfy the target audience.

The synonyms are very important in a bilingual dictionary because they enrich the semantic information which is invariably looked up by the users. Synonyms in a specialised bilingual dictionary can hint at the range of the meanings of each headword, which would assist the specialist dictionary user to make a good selection. According to Svensén (1993:119), 'using synonyms as definitions saves space and the method is entirely valid when the need for semantic precision is not too great…' This action requires careful selection of synonyms and also the aid of a dictionary, because if a user is looking for details in the semantic field the listing of synonyms may cause ambiguity and confusion. The inclusion strategy to be selected by the lexicographer should promote the quick retrieval of necessary information. One of the functions of a specialised dictionary is to impart scientific knowledge: a full synonym should be the first choice, and other alternatives may follow. It was noticed from the EngXhPC that translators tended to be astute in selecting a synonym to be used in certain contexts. Another crucial category of information that can be identified from the word lists developed from the parallel corpus is the orthographic information, which is discussed below.

4.7 Orthographic information

In Chapter 3, section 3.6.1.2, orthography is defined as the information regarding how words are correctly written and spelt in a language. The inclusion of orthographic information in a dictionary assists the users to look up or verify the spelling of certain words. In this regard, as in other aspects, dictionaries are regarded as reliable reference works. The orthographic information is often consulted for inflexion, pronunciation and word division, as well as homonyms. In the word lists displayed in Figure 4.1, 4.2 and Table 4.2 above, the manner in which words are written and spelt is examined in both the source language and the target language due to the direction of the EngXhPC, that is, English to isiXhosa. The wordlists in English reveal uniformity in orthography. This is due to the standardisation and availability of spell-checkers. Like all languages which are dynamic, isiXhosa which has a standardised form of

writing, has undergone several stages of development by acquiring new vocabulary and spelling rules to accommodate new terminology for foreign words and concepts brought in by technology and science, among other influences. The evidence from the EngXhPC wordlists created by ParaConc in EngXhPC pointed to a number of spelling variants. The examples in section 3.6.1.2 in Chapter 3 are a clear demonstration of how translators vary in the spelling of the translation equivalent of the compound noun 'financial year'. In the various texts the word 'financial' is translated differently, in the sense that from the root/stem derivation, different orthographic forms are discovered. It is found in the current parallel corpus that the variation of spelling in isiXhosa is common in the derivation of new words. Table 4.17 below reveals more examples of spelling variants in compounds across the EngXhPC.

Table 4.17: Variants of compounds extracted from EngXhPC

SL term	**TL variants**
financial year	*unyaka-mali, unyakamali, unyaka wemali*
financial report	*ingxelo-mali, ingxelo yemali, ingxelomali*
Amount	*isixa-mali, isixamali, isixa mali, isixa semali*
Budget	*uhlahlo lwabiwo-mali, uhlahlo-lwabiwo-mali*
Retirement	*umhlalaphantsi, umhlala-phantsi*

From the table above, the term amount is spelt differently by different translators as *isixa-mali, isixamali* and *isixa semali*. The concordance lines below demonstrate how these variants can be identified in KWIC.

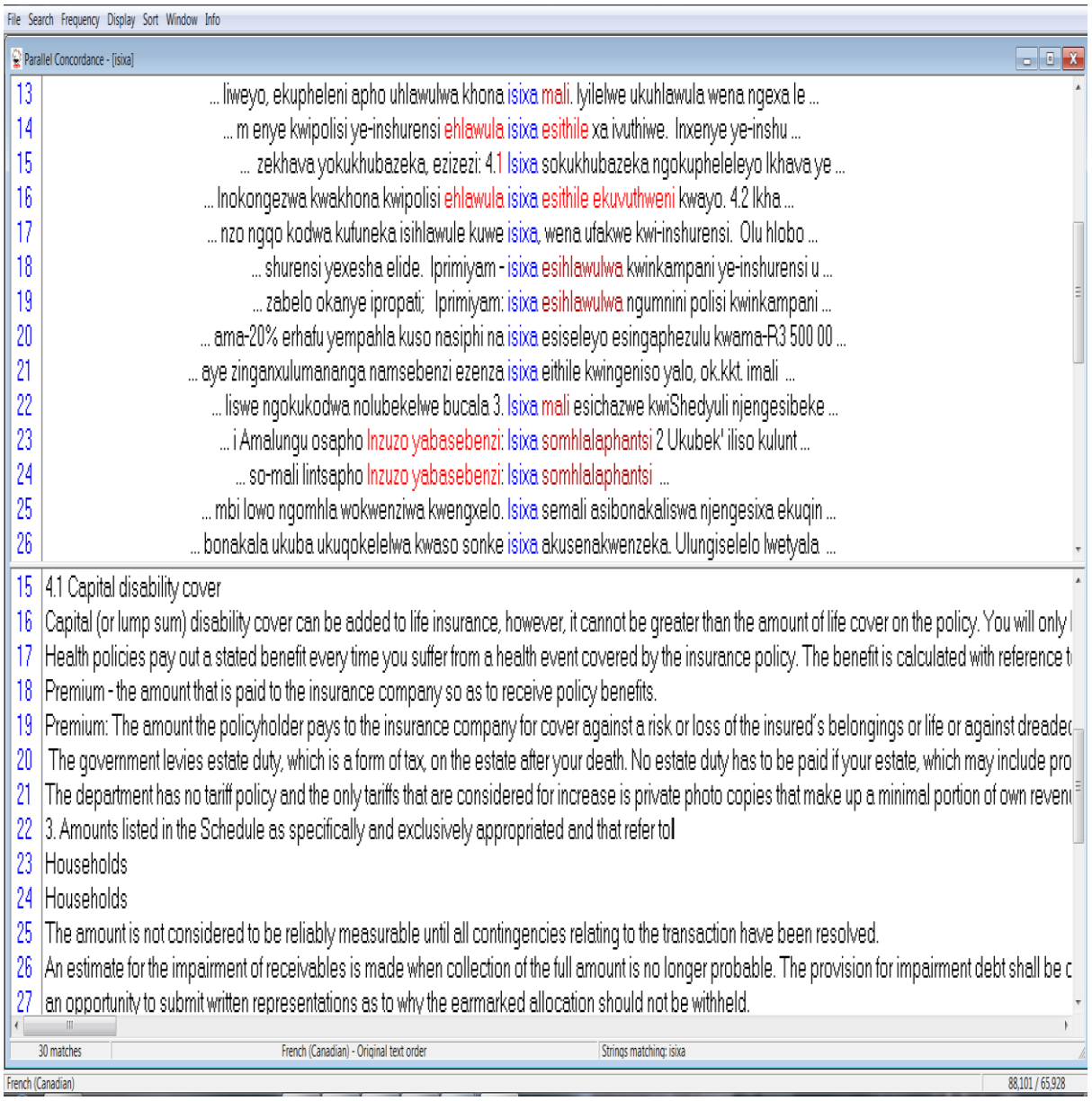


Figure 4.5*: Isixal*amount concordance lines

The screenshot displayed above provides the details of how the financial term, **amount** is translated, and how the various translators have spelt the target equivalents in different styles, thus creating a variety of spelling variants. In lines 13 and 22 of the screenshot, the target equivalent provided is *isixa* (amount), whereas line 23 exhibits *isixa mali*, a very unfamiliar style of compounding where the two elements of the compound noun are separated by a space instead of a hyphen. Line 26 shows *isixa semali* being used with the possessive concord, se-. *Isixa* is frequent in the text with 30 matches.

IsiXhosa has a standardised orthography which was consulted during analysis to compare the results of ParaConc. According to the latest orthography and spellings rules, the compounds are written with a hyphen as in *unyaka-mali* because the latter is derived from two nouns, *unyaka* and *imali* and *isixa-mali* from *isixa* and *imali.* The prefix i- of *imali* is omitted. During the analysis and examination of word lists and cotexts, it was revealed that some of the spelling variants are caused by the influence of English orthography, lack of linguistic competence and research skills amongst translators. These inconsistencies displayed in the corpus data can be standardised and included in the dictionary to provide authoritative guidance to the users of specialist texts. For example, when following the standardised orthography rules and the strings of matches found in EngXhPC, the hyphenated compound is common but certain words are not hyphenated at all. A look at the following screenshot will indicate how 'pension/retirement' is translated in the parallel corpus. In the concordance information of *'umhlalaphantsi'*, the upper window displays the information in isiXhosa while the lower window exhibits the source text information. The source words pension/retirement (see concordance lines 4 and 18 respectively) are consistently translated as *umhlalaphantsi* with no hyphen, whilst the following shows a variant *umhlala-phantsi* which has only 2 matches in the corpus.

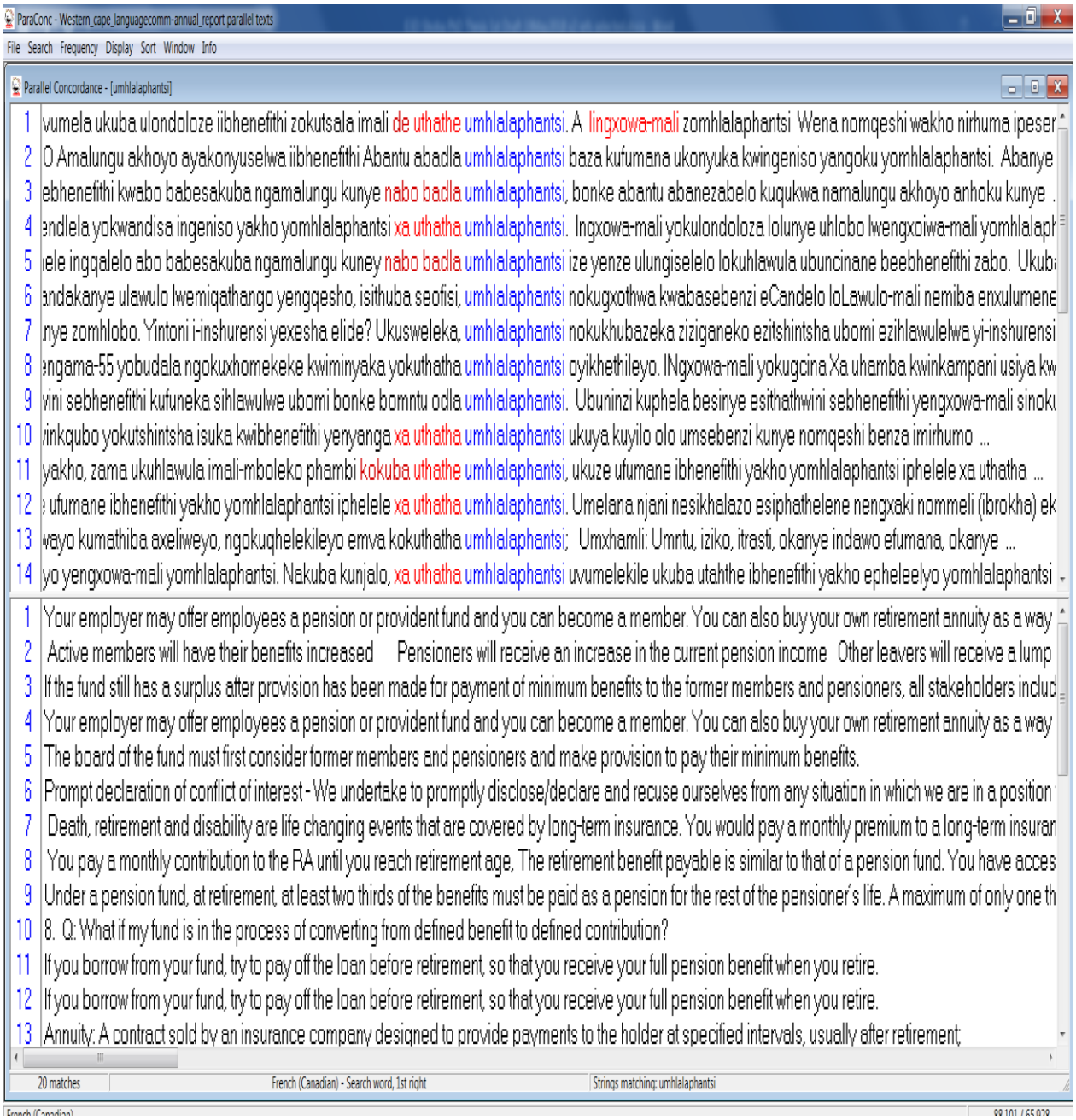


Figure 4.6: Umhlalaphantsi concordance lines

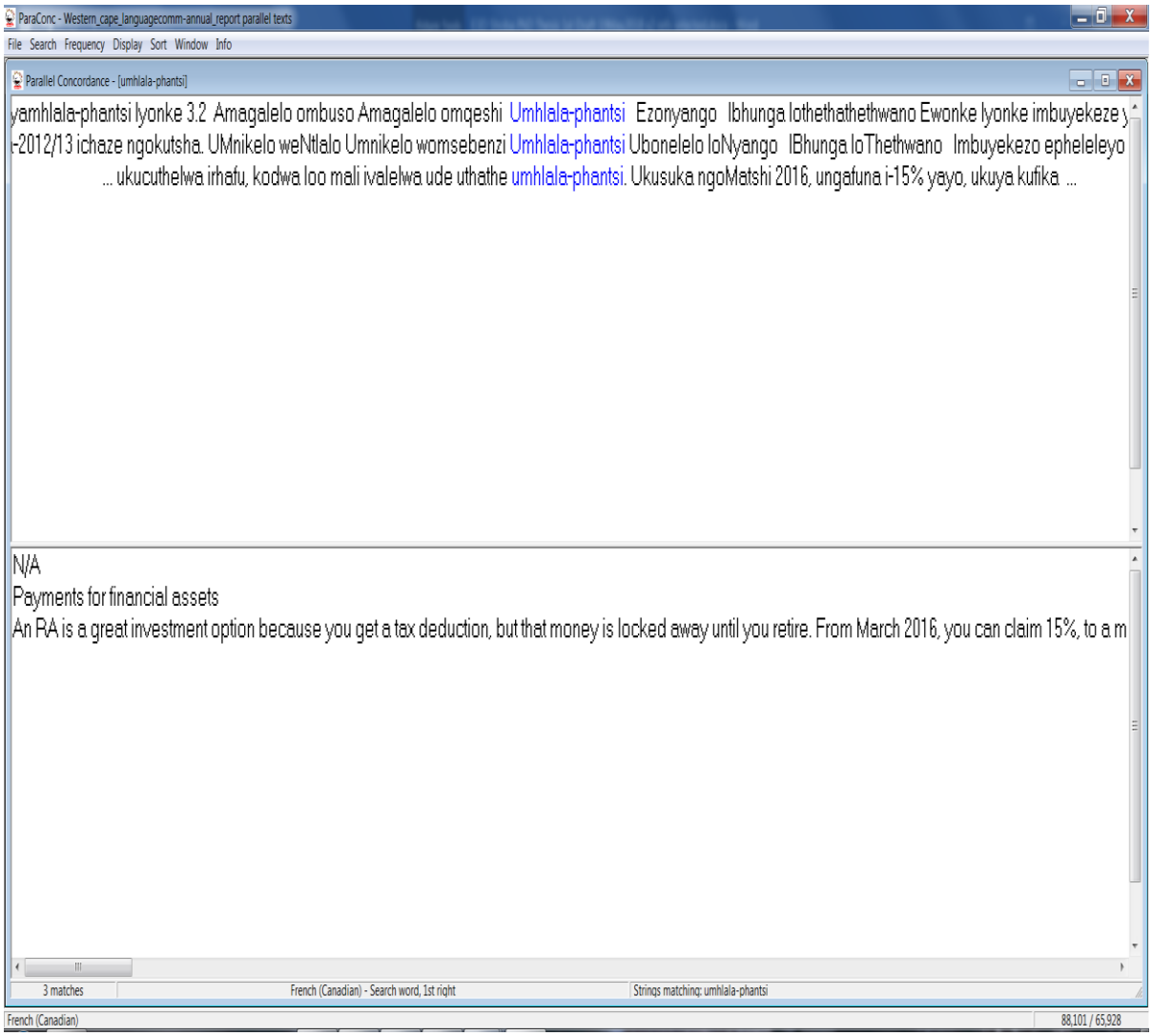


Figure 4.7: Umhlala-phantsi concordance lines

The use of the hyphen in the *Terminology and Orthography No. 3* (1972) booklet was optional in the compounds and duplicated stems, whilst the use of the hyphen in compounds of more than one stem is emphasised in the current orthography. The frequency of hyphenated compounds in Figures 4.6 and 4.7 affords evidence of the two standard spelling rules. The non-use of a hyphen in Figure 4.6 indicates the extent to which the translators abide by the spelling rules stipulated in the old orthography. This evidence challenges a lexicographer to take correct decisions, but fortunately the manipulation of ParaConc through frequency counts is able to provide the most frequently used spelling conventions that can be used as primary forms of the translation equivalent. In addition to the corpus, lexicographers may need to align their work with the spelling rules of the language. This means lexicographers should have strong research skills to complement the evidence from the corpus.

Svensén (2009) attests that frequent spelling variants should be included in a dictionary. The inclusion criterion may also be informed by the kind of the dictionary to be compiled. The information from EngXhPC indicates that numerous spelling variants are also common in loan words. However, few examples of variants are identifiable in single words, such as nouns, verbs, adjectives and so on. Note these few instances from the various texts of the corpus:

1. swipe (verb) = *ukuswayi**ph**a, ukuswayi**p**a*
2. rand (noun) = *ira**nd**i, ira**nt**i*
3. asset (noun) = *iase**th**i, iase**t**i, i-asethi*
4. broker (noun) = ***ibhr**okha, **ibr**okha*,
5. insurance (noun) = *i**nshu**rensi, i**nsho**rensi*.

The source language words above were translated by using borrowing and adaptation into the Xhosa language, or in simple terms, these are indigenised loanwords that have to suit the orthography and the syllable (CVCV) structure of the target language. In example 1 for instance, the last syllable is represented either by pha or pa. When articulated it may sound as either an aspirated */ph/ or* ejective plosive */p/,* hence the variation in orthographic representation. In the second example, the spelling of the last syllable varies in the two words. In *irandi,* the English spelling is retained whilst in *iranti* the last syllable is articulated ***as** /nt'/* The source word asset in Example 3 shows three different spelling variations, that is*, iasethi, iaseti* and *i-asethi.* The variations are based on two factors: aspiration vs ejection and hyphenation. The hyphen is only used between the two similar syllabic vowels. When the vowels are not the same and pronounced differently, a hyphen is not used as in *iasethi/iaseti* (*Spelling and Orthography Rules for isiXhosa,* 2008:133). In order to standardise these terms, orthography rules and frequency lists can be used.

The indigenisation of loanwords may also result in new consonants in the borrowing language. In Example 5, two equivalents of insurance are noted as *inshorensi* and *inshurens*i, resulting in a syllable structure CCVCV. The spelling variation indicates that the texts in the EngXhPC are translated by various translators who opted for different term creation strategies. Performing basic text searches is very necessary in gathering different data in the parallel corpus. Doing so can assist the dictionary compiler in taking decisions based on concrete corpus results. It is worth noting that a

corpus is never perfect; it may contain some errors which necessitate a consultation of other resources. The decision that is taken by the compiler should assist the user to reach the correct information as quickly as possible. In the case of providing suitable equivalents in the TL, the evidence suggests that frequency information could solve the lexicographer's dilemma; for instance, the EngXhPC proves that a compound with a hyphen is most frequent. The bilingual concordances of the two words, namely, retirement and pension, provide a clear illustration of the spellings in context.

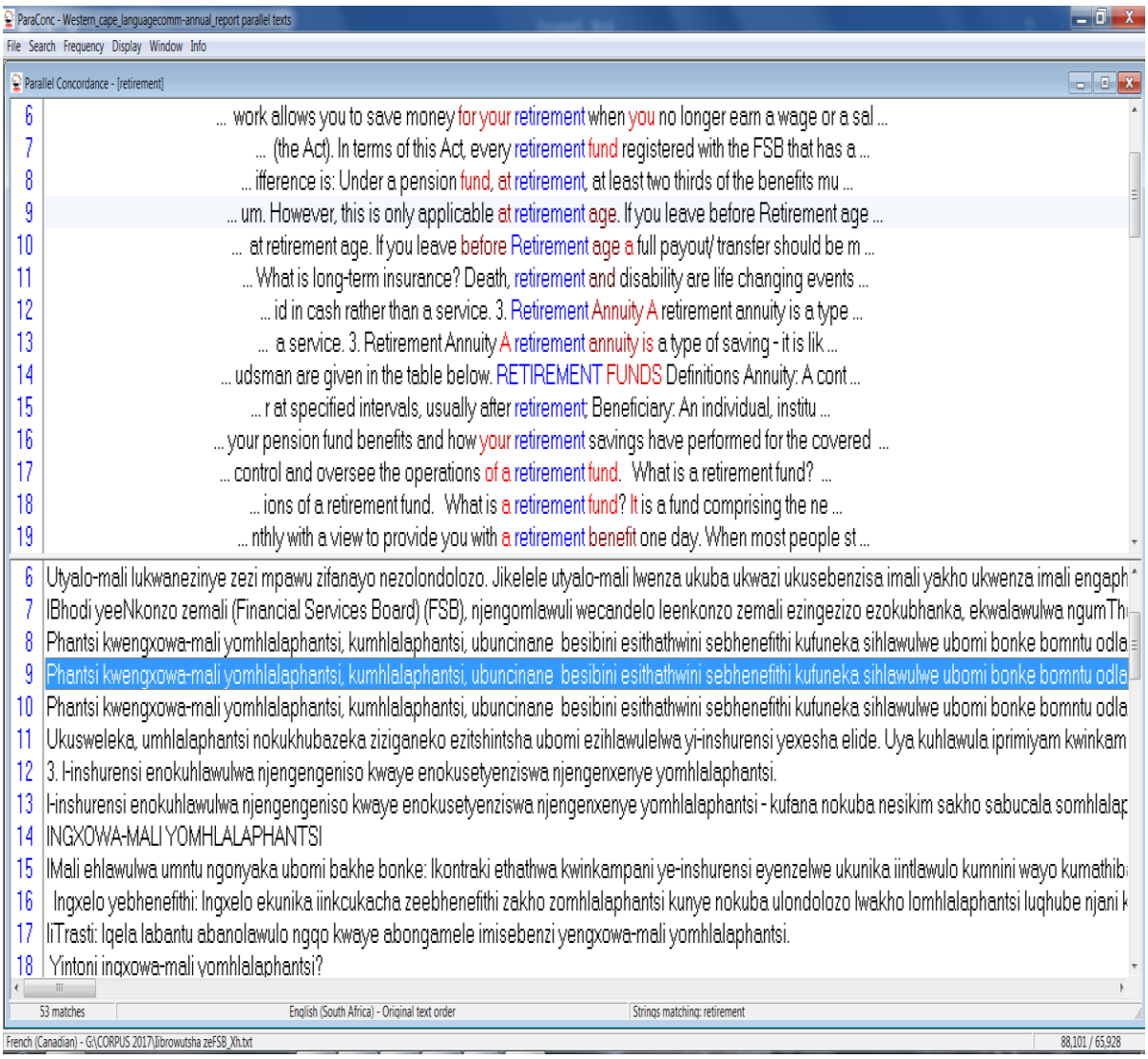


Figure 4.8: Retirement concordance lines

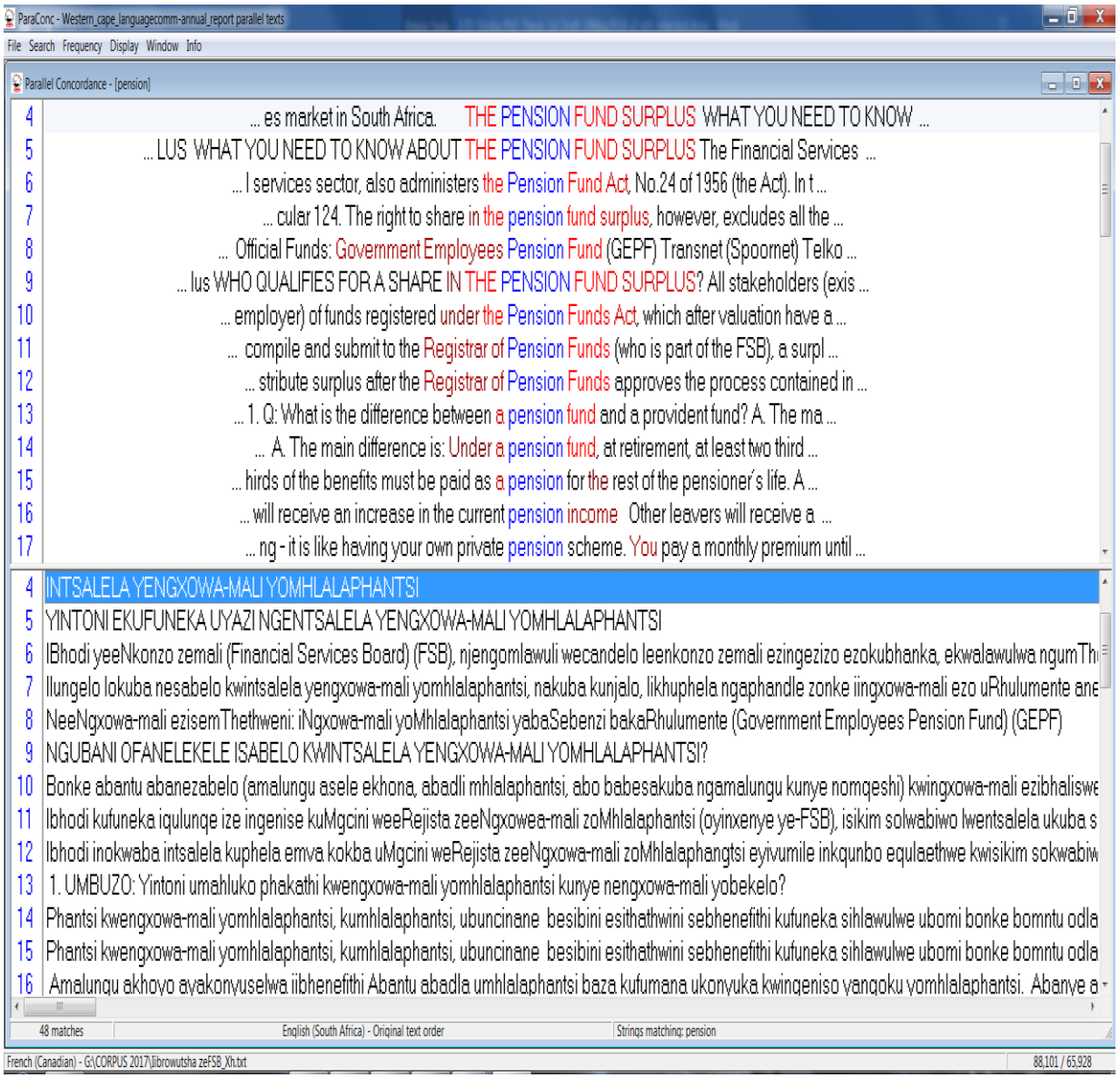


Figure 4.9: Pension concordance lines

The frequency counts by ParaConc can also aid the lexicographer in ordering the two spelling variants: the most frequent one must be listed first while the one with the fewest occurrences should be the last. The frequency tool helps in objective selection of what to include or exclude in a dictionary, a decision that is difficult to take without corpus analysis tools. The frequency also facilitates fast retrieval of the most required information.

4.8 Concordance information

A concordance in Chapter 3, section 3.5.6 is defined as the display of all occurrences of a selected item in a text or corpus. The occurrences as shown above are displayed in KWIC. It will be evident that concordancing is seen as a basic technique in lexicography because it enables the lexicographer to retrieve data that is unique, compared to the conventional methods. There is a consensus amongst scholars that concordances are the primary means of analysis. The parallel concordance lines automated and sorted by ParaConc (cf. Figure 3.8) display various types of corpus data. The quantified corpus data from the EngXhPC provide the contextual information of each search word. Whilst ParaConc performed the quantitative analysis, the researcher relied on the key word in context in order to obtain this contextual information. The lexicographic data that is recognisable in concordances is now discussed under the following subheadings: contextual meaning, collocation and multiword units, and acronyms/abbreviations.

4.8.1 Contextual meaning

Translation equivalents are regarded as the crucial information in any bilingual dictionary. However, it does not help to provide a list of equivalents in a bilingual dictionary without contextual information. The contextual meaning of words is a unique function of ParaConc, gathered through the performance of term searches. The concordances are the basic features of words in context. In the study, this was effectively done and yielded good results from the EngXhPC. The behaviour of search words in the concordances enabled the researcher to pinpoint usage examples that are able to supplement the target equivalents as presented below. The search function of ParaConc is used to search all the words in bold.

Table 4.18: Keywords in context

English source text	IsiXhosa target text
1. Your spending habits play a big role in shaping your **finances.** *(Bona* Magazine*)*	*Indlela osebenzisa ngayo imali idlala indima enkulu ekubumbeni* ***iimali*** *zakho****.***
2. Where do you invest your **money?** (FSB brochures)	*Uyityala phi* ***imali*** *yakho?*
3. Put extra **cash** into a savings account or bond…	*Faka* ***imali*** *egqithisileyo kwi-akhawunti yokonga okanye kwintlawulo yendlu…*
4. If you really need to open a store account, start by drawing up you **budget.**	*Ukuba unyanzelekile nyhani ukuba uvule i-akhawunti, qala ngokwenza* ***ibhajethi****.*
5. Interest: A **payment** made in return to the use of borrowed money.	*Inzala:* ***Intlawulo*** *eyenziweyo ngenxa yokusetyenziswa kwemali ebibolekiwe.*
6. All **surplus payments** are taxable.	*Zonke* ***iintlawulo zentsalela*** *zitsalelwa irhafu.*
7. **Income** is distributed to investors every six months.	***Ingeniso*** *yabelwa abatyali-mali qho kwiinyanga ezintandathu.*
8. You should list the **amounts** for your savings and your investments under your expenses.	*Kufuneka udwelise* ***izixa*** *zolondolozo lwakho kunye notyalo-mali phantsi kweendleko zakho.*
9. Every investment is subject to some form of **risk**.	*Utyalo-mali ngalunye luxhomekeke kuhlobo oluthile* ***lomngcipheko****.*
10. **Irregular expenditure** is recognised as expenditure in the statement of financial position.	*Inkcitho engafanelekanga ibonakaliswa njengenkcitho kwinkcazo eqhube ngayo* ***imali*****.**

The usage examples in Table 4.18 are extracted from concordances displayed by ParaConc. The sentences from EngXhPC parallel texts illustrate how the concepts ‘finance, money, cash’ are used in real-life situations. This type of information retrieved from the translated texts can ease the cumbersome task of inventing examples. Besides demonstrating the syntactical behaviour of the terms, the manner in which they are used can also help in distinguishing the meanings of the three terms.

The contexts in which these words occur confirm that authentic usage examples are retrievable from the parallel corpus. As noted, the advantage of corpus-based usage examples is that they are natural and to some extent more objective than invented examples derived by means of traditional methods.

In section 4.3.2, it was found that the isiXhosa translation equivalent *imali* refers either to finance, cash or money. It is noticed that sometimes translators would use an explanatory equivalent to avoid confusion. For example, the term c*ash* is translated as *imali ezinkozo/imali esesandleni.* Although *cash* has different target equivalents, the equivalent *ikheshi* is more frequent in daily use and in the EngXhPC. This is further confirmed by Fischer et al (2012:85) in the treatment of the lemma cash:

> **cash,** n. *ikheshi**,** imali esesandleni*; (money in any form) *imali*…

In this general-purpose bilingual dictionary, the term is lemmatised and *ikheshi* is listed as the first equivalent. Although example sentences are not provided in this dictionary, the paraphrase and the explanatory note in brackets could provide helpful guidance to the user.

The usage examples selected from Table 4.4.1, when included after each equivalent, may improve access to semantic information. The dictionary users, whether translators or subject specialists, can retrieve the information as quickly as possible. One particular issue in producing specialised bilingual dictionaries relates to the definition and translation of the domain specific usage of ordinary words in specific domains (McEnery and Xiao 2016). A specialised dictionary of specialist terms can benefit from this contextual information to provide more explanations of financial terms. The inclusion of these examples could assist the users in using the lexicographic information in both text production and reception. The corpus-based example sentences are more functional than crafted examples. According to Marza (2009:61), ‘non-corpus-based specialised dictionaries often give improbable, made-up examples using simple language in the illustration of different contexts’. The examples could be used as they are from the above contexts; where they exhibit complexity, a lexicographer may simplify them considering the potential user’s profile. The usage examples may not be necessary in other dictionaries, but in a specialised dictionary where users are looking for the exact meanings, this supplementary information is most significant.

4.8.2 Collocations and multiword units

The definition and classification of multiword units differ, according to various theorists, which results in different terminology. Louw (2006) classifies multiword units into multiword lexical items (idioms), collocations and multiword compound lexical items (with specific reference to compound nouns). The collocations therefore are a type of a multiword unit. They are therefore called multiword units (MWU) or multiword expressions (MWE). In this analysis, the phrase multiword units (MWUs) is consistently used. According to Svensén (1993:101), 'information on collocations is important in both monolingual and active bingual dictionaries, since the user cannot be expected to know which words customarily occur together'. The co-occurrence of words is a common feature in parallel concordances. The manual identification of collocations and multiword units for inclusion in a dictionary is a mammoth task, but with ParaConc, this function is very quick and effective. To retrieve collocations and other multiword units, a search word is typed in the search box and a concordance window is displayed with source language and translations on the lower window. The concordance view of the financial term 'risk*'* reveals all the words that co-occur with the keyword, the node word in the middle and its contexts (see Figure 4.10 below).

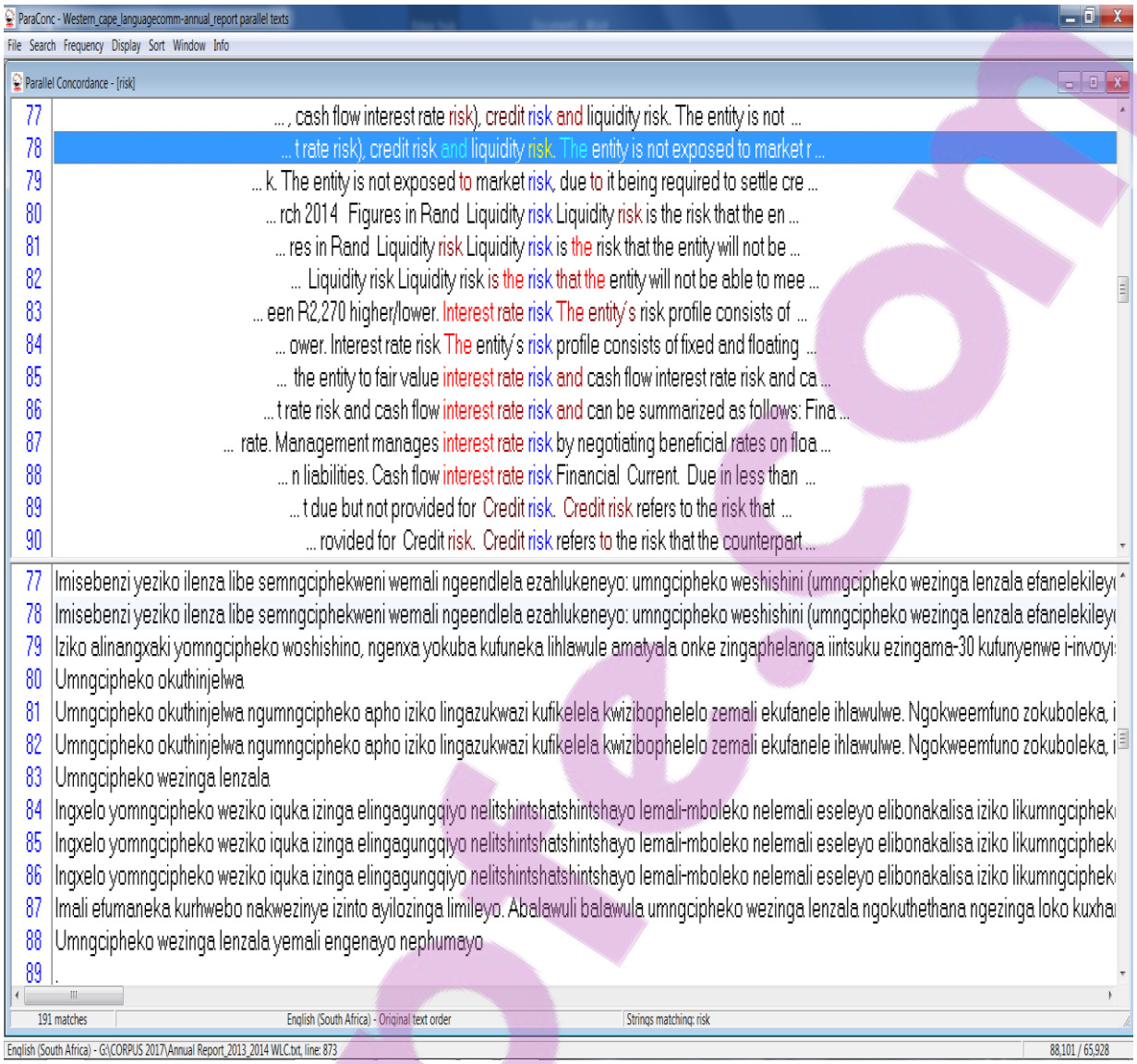


Figure 4.10: A concordance of risk

The screenshot in Figure 4.10 depicts the KWIC in the source language, followed by translations in isiXhosa as the target language. In studying and analysing it horizontally, the words that co-occur with 'risk' in the concordance are classified as follows:

Example 1: Function word collocates:

i) of risk

ii) risk due to

iii) per risk

iv) risk of

v) and risk/risk and

Example 2: Adjectival and adverbial collocates

i) strategic risk

ii) significant risk

Example 3: Nominal collocates

i) **credit** risk

ii) risk **management**

iii) risk **register**

iv) **liquidity risk**

v) **Interest rate risk**

The above examples demonstrate the different word categories that might frequently keep the company of the word ‘risk’ in the EngXhPC. These are either grammatical words, as in example 1 and 2, or lexical or content words, as in example 3. The grammatical words do not have meaning when in isolation and need to be used with content words, thus providing unique information. The grammatical collocates when used in context in a dictionary should assist the users when writing. The words that always accompany each other will be reproduced using adequate collocates. The inclusion of collocates plays a significant role in dictionary entries, because they disclose words that always go together. They also aid in identifying idioms and phrases in the contexts. The collocations extracted in the parallel corpus can be included in the microstructure to empower the users in encoding functions.

The horizontal lines in the concordance in the upper and lower window also made it easy for the analyst to read the strings of various combinations, thus identifying lexical items that are combined to form one meaning. Table 4. 19 tabulates the top 10 collocations identified from the concordance view in Figure 4.10 above, through qualitative analysis by the researcher.

Table 4.19: Identification of collocations from a concordance view of the EngXhPC

English collocations	**IsiXhosa translations**
Risk management	*ulawulo lomngcipheko*
Enterprise Risk Management	*iButho loLawulo Lwemingcipheko [sic]*
Fraud and risk management	*ubuqhetseba nolawulo lwemingcipheko [sic]*
Risk register	*irejista yemingcipheko*
Interest rate risk	*Umngcipheko wezinga lenzala*
Credit risk	*umngcipheko wamatyala*
Risk profile	*inkcukacha zemingcipheko*
Market risk	*umngcipheko wemarike*

The collocations in the concordance, and in the column, are sorted according to the frequency of co-occurrences. The equivalent translations are picked up from the lower window of the concordance. The sentences depict the contexts of the strings in real use and communication in the target language. The context and collocates make it straightforward for the analyst to determine the meanings of words: 'Words build meaning in context, and meaning is the result of the association of words' (Pensec 2016:20). This is true, particularly in the specialised texts, because some words may sound and look like general words, but context will clearly draw the line between LSP and LGP terminology. The choice of translation equivalents in a corpus should be thoroughly examined since some words may be mistranslated. For example, in Table 4.19, the first two collocations, 'enterprise risk management and fraud and risk management' show mistranslations. In the source language, 'management of risk' is the core of the expression, meaning that isiXhosa translation equivalent should have begun with *'ulawulo'* in order to be accurately translated.

These examples emphasise the point made earlier in subsection 4.3.4.3 that corpus evidence is not always foolproof.

The list of multiword units or collocations identified from the EngXhPC can be used in selecting the candidate headwords in a specialised dictionary. The common practice in specialised lexicography is to lemmatise single words and multiword units or their collocations in alphabetic order. In *the Oxford Dictionary of Finance and Banking* 2014:402-404), the headword 'risk' is lemmatised and followed by all combinations of *risk*: for instance, following the noun 'risk', come risk-adjusted assets, risk-adjusted discount rate, etc. These multi-words have 'risk' as the central word but they predict one word with one meaning. More examples of multiword units or compounds and their translations are discussed below.

Table 4.20: Compounds/multiword units identified in EngXhPC

1.	Retirement fund	*Ingxowa-mali yomhlaphantsi*
2.	Retirement annuity	*Intlawulo-mali yomhlalaphantsi*
3.	Retirement benefit	*Inzuzo yomhlalaphantsi*
4.	Accounting officer	*Igosa elinoxanduva lokuphendula*
5.	Contingency fund	*Ingxowa-mali yemini kaxakeka*
6.	Auditor General	*Umphicothi-Zincwadi Jikelele*
7.	Financial management	*Ulawulo lwezemali*
8.	Financial resources	*Imithombo yezemali*
8.	Collective investment scheme	*Isikim Sotyalo-mali Seqela*
9.	Pension fund surplus	*Intsalela yengxowa-mali yomhlalaphantsi*
10.	Government Employees Pension Fund (GEPF)	*Ingxowa-mali yoMhlaphantsi yabaSebenzi bakaRhulumente*

Table 4.20 above is an illustration of collocations and compounds that are displayed in the concordance through a key word search as shown previously. ParaConc has revealed that hyphenated compounds are the properties of the frequency list, whilst open compounds collocate with search words. Open compounds in Table 4.20 above are a combination of two or more lexical units; hence they are classified as multiwords.

In examples 1, 2, and 3, retirement fund, retirement annuity and retirement benefit are the results of the search term 'retirement' from the parallel corpus. All the three examples co-occur on the right of the search term and have different meanings; therefore, they are the headwords in their own right. Retirement, for example, means money that is paid to any person who has reached a pensionable age, while a retirement fund is a lump sum of money reserved for retirement. The information that is retrievable from the corpus is useful in a dictionary as it provides more understanding of these specialist terms which may be new to the general user. A target reader who is a layperson or non-specialist could be more financially empowered to access the information in the terminology that is used in daily life. In relation to the SL compound 'retirement fund', the use of *ingxowa-mali* adds more meaning to the compound, if one is aware that *ingxowa* in isiXhosa is a hessian bag traditionally used as a container for maize meal (it literally means money pool). A word that is known in isiXhosa has been attached to *imali,* an LSP word, in order to create a new word. The translator has used a semantic transfer or expansion strategy. A hessian bag that used to be a container of maize has been shifted in meaning to be a pool of money. As it is, the word has attained a specialised meaning. The meaning of multiword units lies with all their components. Semantic expansion refers to a strategy where a term that already exists in a language becomes attached to another term, and acquires an additional meaning (Mtintsilana and Morris, 1988). This is indicative of innovative translation strategies that contribute to the expansion of terminology in isiXhosa.

All the words in the two tables are word combinations which are a result of word compounding. In example 9, the source language compound "pension fund surplus' is derived from the combination of three nouns: pension + fund + surplus = *intsalela+ingxowa-mali+ umhlalaphantsi*. A translation equivalent is a paraphrase or description: *intsalela yengxowa-mali yomhlalaphantsi*. Because of syntactic rules and concordial agreements the possesive concords *ye-* and *yo-* are prefixed to *ingxowa-*

mali and *umhlalaphantsi* respectively. The collocations extracted by ParaConc also represented phrasal verbs and idiomatic expressions; these are presented below.

4.8.3 Extracting phrasal verbs and idiomatic expressions

The technical language or LSP is also characterised by idiomatic expressions and phrases which also fall under the subject being discussed. Through the analysis of bilingual concordances, idiomatic phrases are easily identified. The function words that are listed in the frequency lists and also displayed in the concordances might form some idiomatic expression or phrases. Below is a concordance of a search word, 'debt', and consequent finance-related expressions or phrases.

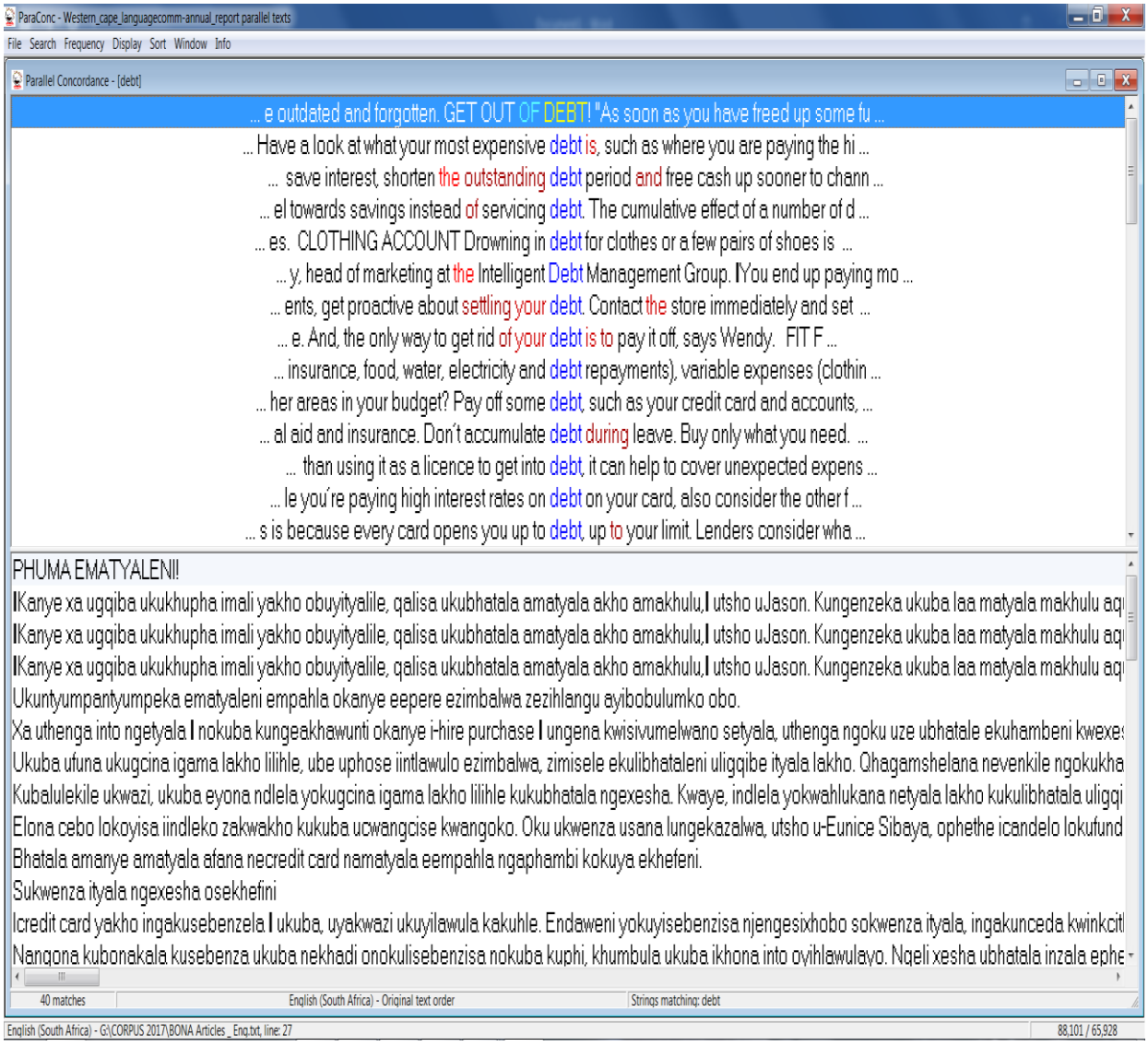


Figure 4.11: Collocations and expressions of the word debt

Example 1: Get rid of debt, pay off

SL: And the only way to **get rid of your debt** is to it **pay it off.**

TT: *Kwaye indlela* ***yokwahlukana*** *netyala* ***kukulibhatala uliggqibe***.

Example 2: Out of debt

SL: Get out of debt.

TT: *Phuma ematyaleni*.

Example 3: Drowning in debt

SL: Drowning in debt for clothes or a few pairs of shoes is not wise.

TT: ***Ukuntyumpantyumpeka ematyaleni*** *empahla okanye eepere ezimbalwa zezihlangu ayibobulumko*.

Example 4: Settle a debt

SL: If you want to keep your credit score in good shape, and you've missed a few payments, get proactive about **settling your debt**.

TT: *Ukuba ufuna ukugcina igama lakho lilihle, ube uphose iintlawulo ezimbalwa, zimisele* ***ekulibhataleni uligqibe ityala*** *lakho*.

The above illustrations indicate finance-related sayings or expressions attached to the search word 'debt', how they are translated into isiXhosa and how they behave in context by following syntactic rules of the language. For example, 'drowning' in general language means to be under the water, but when used with the word debt it becomes a specialised idiomatic expression meaning to have huge debts. 'Drowning in debt' as indicated in Example 3, is translated as *ukuntyumpantyumpe*ka *ematyaleni*. The selection of *ukuntyumpantyumpeka* as its equivalent in isiXhosa shows how the translator has adapted it to the target culture. In the *GDX* (Volume 2 K-P), *ukuntyumpa-ntyumpeka* means 'to wallow in the mud; of one who is on the verge of sinking; struggle to save oneself; and struggle under the strain of difficulties or debts'. Drowning in debt can also be translated as *ukweyela ematyaleni*, but the translator has chosen *ukuntyumpantyumpeka* in order to send a strong message to the audience. This is illustrative of cultural embedding in financial texts (see section 2.2.3 in Chapter 2). The selection of this term has bridged the gap between source language and the target language culture. Another observation from the above contexts is the synonymy between expressions. In examples 1 and 4 above, the phrasal verb, pay off, is used

synonymously with ‘settle’ and translated as –*bhatala uligqibe ityala*. Such data would be difficult for the lexicographer to collect manually.

The identification of these expressions from the parallel corpus requires a linguistics background and subject specific knowledge. As discussed, the identification of multiword units that are closely attached to the search word is a complex and time-consuming process, but the exploitation of the corpus tool has made it easier and more convenient. The retrieval of these lexical items also requires human intervention. Hence the researcher has scanned and studied the cotexts. The inclusion of multiword expressions harvested from concordances can significantly improve the dictionary entries of African languages (Otlogetswe 2007:448). The above analysis demonstrates that if such information can be included in a dictionary the users will more easily understand the specialised terminology. The search for multiword units and collocations is characteristic of corpus linguistic methods. This information further confirms that the utilisation of KWIC information from the EngXhPC can aid a lexicographer in identifying different kinds of multiword units and give guidance on the lemmatisation strategy to be followed in a specialised dictionary. As observed from parallel concordances and the referenced LSP dictionary above, multiword units, particularly open compounds, form an integral part of the macrostructure in specialised lexicography. In some dictionaries multiword units are entered as sub-entries at a microstructural level of the entry. The *Oxford English-Xhosa Dictionary* was consulted to verify the inclusion and treatment of the key word ‘debt’. The target equivalent is provided with additional information. The phrasal verb ‘get out of debt’ is included in the micro-structure with its equivalent *phuma ematyaleni*. This also indicates that translators have consulted other reference works during the translation processes.

According to Pritchard (2015) multiword lexical units are a typical feature of specialised dictionaries. The abundance of MWUs in the EngXhPC, particularly compounds, provides evidence that translators have used compounding in order to convey a precise message to the target audience. This is also an indication that compounding is highly productive in specialised translation. ‘Data on collocations provides microsyntactic empowerment, especially to users employing their dictionary in an encoding task’ (Louw, 2006:99). The inclusion of multiword units in the macrostructure is the preferred approach in current specialised dictionaries. This gives them full lexicographical treatment and therefore helps the users to collect optimum information

with easy access. Another type of data that is recognised from the concordance view is a range of acronyms in contexts; these are discussed below.

4.8.4 Acronyms and abbreviations

An acronym in the *Oxford Dictionary* is defined as an abbreviation formed from the initial letters of other words and pronounced as a word. These letters conform to the orthographic rules of isiXhosa: ‘Abbreviations are always pronounced as a sequence of letters’ (Moropa, 2005:172). The pronunciation is a distinctive factor. JSE for example is an abbreviated form of Johannesburg Stock Exchange but is not uttered as one word, while SARS is the acronym of South African Revenue Services since it is pronounced as a single word. The distinction between the two is not always clear in the written form: therefore, the two words, acronyms and abbreviations, will be used interchangeably in this analysis. Acronyms form an important category of information in the specialised dictionaries. Although the frequency lists give clues to the number of acronyms that are frequent in the EngXhPC, the investigation of concordance lines revealed the acronyms in their immediate contexts in both English and isiXhosa. In Figure 4.10 above, in studying the information in the upper window, the following English abbreviations were easily spotted:

1. Enterprise Risk Management (ERM)
2. Public Finance Management Act (PFMA).

In the target language the name from where the acronym was derived can be translated but the actual acronym is not altered, though its behaviour in the sentence is adapted to that of isiXhosa. Table 4. 21 records the context syntactical behaviour of the most common acronyms and abbreviations retrieved from the EngXhPC.

Table 4.21 Examples of acronyms and abbreviations from concordances

Source text	Target text
1. The GRAP standards applicable to the entity and adopted as follows:…	***1****. Imigangatho ye-GRAP esebenzayo kwesi sigqeba nelandelwayo ime ngolu hlobo lulandelayo:…*
2. The Audit Committee reports that it has complied with the responsibilities arising from section 51(1)a of the PFMA and Treasury Regulations 27.1.	***2.*** *IKomiti yoPhicotho-zincwadi yenza ingxelo yokuba iye yalufezekisa uxanduva lwayo oluchazwa kwicandelo 51(1) a lomthetho oyi-PFMA nakuMgaqo KaNondyebo 27.1.*
3. The **FAIS** Act is one of the laws administered by the Financial Services Board.	*UmThetho **we-FAIS** ngomnye wemithetho elawulwa yiBhodi yeeNkonzo zeMali.*
4. An **RA** is a great investment option because you get a tax deduction.	***I-RA*** *luhlobo lotyalo-mali olulungileyo kuba ufumana ukucuthelwa irhafu.*
5. This is done by providing information from **PERSAL** (the Personnel Salary System)…	***TT:*** *Oku kwenziwa ngobonelela ngeenkcukacha **zePERSAL** (Inkqubo yokuHlawulwa kwaBasebenzi)…*
6. **ETF's** are more unpredictable than the other two investments**.**	***Ii-ETF*** *azigqibeki njengezinye iintlobo ezimbini zotyalo-mali.*
7. **SCOPA** resolutions.	*Izisombululo **zeSCOPA.***
8. The above funds do not fall under the jurisdiction of the **FSB.**	*Ezi ngxowa-mali zingentla aziweli phantsi kolawulo lwe-**FSB.***
9. The participatory interest price, also known as unit price or the net asset value **(NAV)**, depends on the market value of the underlying investment in which proof of money is invested.	*Ixabiso lesibhambathiso sezabelo ekwaziwa njengexabiso leyunithi okanye ixabiso le-asethi eliseleyo (net asset value) **NAV** lixhomekeke kwixabiso lemarike lotyalo-mali elisele likhona kwimali ekutyalo-mali.*
10. Contravening or failing to comply with **FMA** rules and directives of the **SRO.**	*Ukophula okanye ukusilela ukuthobela i-**FMA** kunye nemiyalelo ye-**SRO**.*

The KWIC information above demonstrates the use of acronyms when referring to laws and regulations. The context in which the word is used in English allowed the translator to borrow it as it is. The context therefore can tell whether a loanword is appropriate or not. There are two types of information that are discovered from the above sentences, namely the meanings of the acronym and the orthographic form. The translated acronyms in the lower window of the concordance view may supply valuable information on appropriate equivalents within real context and use. In the above examples, the acronyms in isiXhosa are preceded by prefixes, whilst the English stem is retained. However, inconsistent spelling of certain acronyms in different corpus texts was noted. According to the updated *Orthography and Spelling rules for isiXhosa (PanSALB 2008),* 'a hyphen is not used between an isiXhosa prefix and foreign-language word, proper name, or acronyms'. In the last example, *zePERSAL,* there is no hyphen between the prefix and the English acronym. The translator followed the new orthography rules. The majority of examples, the English acronym and the prefix, are separated by a hyphen, contrary to the revised rules. It is noted that some translators used the old system of writing; therefore, lexicographers should use the corpus data with caution because parallel corpus built from 'real life' texts might be flawed at times. In order to construct a representative and factual dictionary, there is need at times to go beyond the parallel corpus. This is because, as indicated, dictionary users always expect to obtain the correct information from a dictionary. In a case of spelling variants, it is suggested here that the standard orthography should be the only choice. In Example 9, the abbreviation NAV is preceded by its full form, 'net asset value' which is then translated as *ixabiso le-asethi eliseleyo* in isiXhosa. Because of the translation the abbreviation keeps its source language form with no prefix. The data in EngXhPC consequently reveals that some acronyms are not always translatable. The translator's style in example 9 might be beneficial even in a dictionary entry if a paraphrase like this always accompanies some complicated or new acronyms, for the benefit of the users. The concordances in the EngXhPC yielded useful results in terms of bilingual terminology extraction. Parallel texts are the best option for data mining because the texts were translated for a target audience that could be the potential users of the specialised dictionary.

Although the texts were translated by different translators, to a large extent the terminology used is appropriate to be entered in a bilingual specialised dictionary of finance. Some fictitious example entries follow.

1. **money** (n) - *imali*
   Ex: *Uyityala phi **imali** yakho*? (Where do you invest your money?)

2. **account** (n) – ityala, *iakhawunti, i-akhawunti.*
   Ex: Ukuthenga ngetyala/ngeakhawunti asibobulumko. (To buy on credit is not wise).

3. **pension** (n) *umhlalaphantsi, ipenshele*.
   Ex: *Abantu abaninzi bathatha umhlalaphantsi phambi kwexesha* (Many people take early pension).

4. **pension fund** (n) *ingxowa-mali yomhlalaphantsi.*
   Ex: *Yintoni umahluko phakathi kwengxowa-mali yomhlala-phantsi nengxowa-mali yobekelo?* (What is the difference between a pension fund and provident fund?).

5. **NAV (abbr.)** *ixabiso le-asethi eliseleyo* **(iNAV).**
   Ex: Ixabiso le-asethi eliseleyo (NAV) lixhomekeke kwixabiso lemarike yotyalo-mali. (The net asset value (NAV), depends on the market value of the underlying investment).

These are dummy entries that could form a bilingual specialised dictionary entry that would suit the variety of users of both English and isiXhosa. The aim of this attempt is not to prescribe to the lexicographer but rather to demonstrate that bilingual terminology extracted from the EngXhPC is able to provide various information categories that would constitute a bilingual dictionary entry of finance.

The indispensable parts of the entry word in an envisaged bilingual dictionary as demonstrated in the few examples provided in this chapter could offer the following lexicographic information:

- A headword in its canonical form
- Translation equivalents in their basic form and synonyms
- Alternative form of spellings (cf. example 2)
- Grammatical information which is optional in specialised dictionaries
- Multiword units and collocations
- Contextual information in the form of example sentences.

The parts of speech labels may not be included in a specialised dictionary, because users generally consult such a dictionary for details about domain specific information. An entry with the types of information listed above follows a user-oriented approach proposed by Gómez Gonzálev (2005) for a specialised bilingual dictionary. The macrostructure has different headwords, namely, single words, compounds, acronyms and abbreviations. The illustrative examples as in example 3 above not only show how the word can be used, but also give an indication of words that often co-occur with headwords and hint at the meaning. For example, the word pension customarily goes with the verb 'take' in English and *ukuthatha/ukudla umhlalaphantsi* in isiXhosa. This wealth of information in the parallel corpus could improve the dictionary content. Prinsloo (2015) observes that the tendency of bilingual dictionaries only to include translation equivalents does not fulfil the most basic text reception (decoding) needs of the users and is not useful for text production (encoding purposes). The data types extracted by the corpus tools in this EngXhPC are able to address the inadequacies of bilingual dictionaries in general and in particular bilingual specialised dictionaries in isiXhosa. The corpus data above complement the intuitive ability of a human being; the lexicographer does not have to translate any longer. The translations as they stand are functional with a choice of functional equivalents. The following section contains a discussion on the two main lexicographic functions aimed at in this study.

4.9 Dictionary functions

According to the functional theory of lexicography already discussed in Chapter 2, Section 2.8.2, users consult dictionaries in cognitive and communicative situations. The dictionary functions that can be fulfilled by such dictionary entries are cognitive and communicative ones (cf. Fathi, 2014; Nkomo, 2008; Tarp, 2008; Bergenholtz and Nielsen, 2006). The cognitive functions assist the users in knowledge acquisition, whilst communication-oriented functions provide problem-solving information in text production and text reception. The two types of functions are linked to the user's needs and the type of information that is included in the dictionary. The type of data that is extracted from the EngXhPC may be helpful in achieving these two functions. The information provided in the sample entries above could assist the users for example to (i) understand the finance-related terms which are very often complicated, (ii) use them in different situations and contexts, (iii) provide information about the language LSP and LGP linguistics, (iv) translate texts from English into isiXhosa.

The reality in South African lexicography, as noted, is that bilingual dictionaries are used by different users for different reasons; therefore the inclusion of comprehensive dictionaries entries may improve the situation. In isiXhosa for instance the *English-Xhosa Dictionary* is widely used by translators, learners, researchers, English and isiXhosa speakers, so that it is still regarded as a useful resource in isiXhosa. The kind of lexicographic information provided in the above mentioned dictionary is the reason behind its being preferred. The overview of isiXhosa lexicography in Chapter 2 reveals that for decades, bilingual dictionaries have been regarded as tools for terminology development and standardisation.

4.10 Challenges and weaknesses of extracting corpus data

During the manipulation of data in the parallel corpus, some challenges were experienced. It is also worth mentioning that the challenges and weaknesses observed during the extraction of bilingual information are general. The data from translated texts contained a number of errors, some of which have already been discussed in the previous sections of this study. Thus, lexicographers have to clean the corpus thoroughly before use because what they put into a corpus is what users get out. The translators of some texts showed lack of experience in translating, and disregarded the

audience they were writing for, which might have been due to a lack of translation briefs from the clients. Nord (1997) describes a translation brief as an instruction given to the translator which states the purpose and the type of the audience the translation is aimed at. When a clear brief is provided, a translator is able to select suitable words. Lexicographers, therefore, need to verify information by checking other sources. This could improve the quality of dictionaries. In relation to translation studies specifically, this study points to the fact that there is a dire need to train translators because they play a crucial role in meeting the communication needs of many. Translators should take their role seriously and carry out their work professionally. The authors and copyright owners should also be conscious that the information posted for public consumption is scrutinised before any publication. Corpus data therefore should be used cautiously in order to retain the valuable characteristics they possess, namely, reliability and authenticity.

4.11 Conclusion

In addressing the two main research questions stated in Chapter 1 of this study, this chapter analysed, interpreted and discussed the findings. The parallel corpus constituted by the source texts and their translations was interrogated by means of ParaConc. The pre-alignment of parallel texts resulted in ensuring an accurate correspondence of the SL and TL pairs, which made it easier for ParaConc to analyse the parallel corpus. The amount and quality of data extracted by this concordancer has demonstrated that corpus-based lexicography is unavoidable in this digital era. The corpus tools that are found to be the core of both the quantitative and qualitative analyses of the given data are the frequency lists and the concordances. By exploiting these two basic tools, various data types suitable for a bilingual dictionary of finance were easily retrieved from the EngXhPC.

To reiterate: the first initial step of dictionary making is the compilation of word lists. In traditional lexicographic practice, the lemma lists are compiled manually and the process is sometimes marred by many errors, because alphabetisation is also manual. The frequency lists that are generated by ParaConc are automated, providing frequency counts of each word. The listing of all the words in the word lists has been found to be an effective way of selecting headwords according to the alphabet and

frequencies. The frequencies also assisted in performing more searches, guided by the search items. This revealed the centrality of the key words in exploiting the parallel corpus. The following data types are the outputs of the frequency lists:

- Inventory of word forms in EngXhPC
- Display of spelling variations
- Inflections
- Word divisions
- Frequently used translation equivalents and the like.

It is noticed that the kind of information that is extracted by the frequency command contributes more to the macrostructure of the dictionary, with the headwords as heads of dictionary entries. The search words are the entry points to the various queries in the parallel corpus, just like the headwords in the dictionary entry. The concordance is proven to be the centre of corpus analysis. The studying of the concordance lines provides a wealth of information that is not easily found in other types of texts. Through a basic search, ParaConc displayed concordances with KWIC. The KWIC format allows for quantitative and qualitative analysis. The type of information that is retrievable through the concordances and KWIC consists of:

- Translations in parallel
- The typical collocates of the search item
- Words in contexts, meaning in use
- Syntactic and collocational behaviour of words.

The analysis of the information in the concordance provided the different kinds of translation equivalents, as well as a variety of multiword units such as compounds, collocations and idiomatic expressions. This chapter has also revealed that authentic usage examples are extracted from the concordances. Another observation made is the predominance of compounds in the specialised language.

For instance, it is fascinating to observe how many nominal compounds with the word *imali* are extracted from the EngXhPC: *utyalo-mali, ingxowa-mali, ulwabiwo-mali, ucwangciso-mali, ubalo-mali, isixa-mali,* etc. This indicates that the parallel corpus is subject-specific and is a repository of complex LSP vocabulary.

In the analysis of equivalents extracted by ParaConc in the EngXhPC, the interesting findings on term formation processes disclosed what was taking place during the process of translation. In attempts at coming up with appropriate and target-oriented translation equivalents, translators selected various translation strategies. The extraction of idiomatic expressions in the parallel corpus demonstrates the qualities of a technical language. In section 4.8.3, the concordance of the source word ‘debt’ displayed the collocates that formed phrasal expressions when co-occurring. Through ParaConc, EngXhPC effectively revealed the recurrent patterns of multiword units as they occur in language discourse. The frequent occurrence of the verb ‘plummet’ with the verb price constitutes typical financial language that would be difficult to discover without a parallel corpus of natural language. The reason for translating the texts into isiXhosa was to communicate with the target users in the language they understand. Accordingly, the vocabulary is designed for practical purposes with the objective of communicating the specialist terminology in the home language. For this reason, the EngXhPC can be used as the primary source of authentic dictionary information.

The conclusion reached in this chapter is that a parallel corpus is a rich source of authentic data as already mentioned. The linguistic investigations of the corpus texts would not be feasible without corpus analysis tools such as frequency lists with statistical information, words in context and the retrieval of complex units such as collocations. The extraction of the latter is recognised as a major challenge in lexicographic practice but ParaConc has made available instant and reliable data.

As evidenced by corpus data in the parallel corpus, the bilingual terminology can be used to improve the future specialised bilingual dictionaries of isiXhosa. The products of translation, the translated texts in the corpus, have demonstrated how the practice of translation is able to influence lexicographic practice. The information also indicates how the corpus-based approach is able to benefit both translation and lexicography. The following chapter concludes this study.

CHAPTER 5

Conclusion

5.1 Introduction

This concluding chapter provides a restatement of the research problem and aims and an overview of the chapters. The proof that the aims that were set out have been achieved is presented under the summary of findings. This is followed by an assessment of the contributions of the study to the body of knowledge. Various recommendations are made, reflections on the limitations are discussed and finally, suggestions for future research are proposed.

5.2 Aims and research questions

The main aim of the current study was to investigate how specialised parallel corpora can be used in the compilation of bilingual dictionaries of financial terms involving English and isiXhosa. In order to achieve this, a specialised parallel corpus composed of translated texts was built. The secondary aims were to extract bilingual lexicographic information that could be included in a specialised bilingual dictionary. A corpus analysis tool, ParaConc was used to analyse the data from the EngXhPC.

These aims were addressed by answering the two research questions: (1) How can ParaConc be used to extract bilingual information from specialised parallel corpora? (2) How can information extracted from parallel corpora be used to create bilingual dictionaries?

5.3 Overview of chapters

In setting the scene for this study, an outline of each chapter is presented as follows:
In **chapter 1** an appropriate background and rationale was provided. The research statement was also contextualised. Despite the elevation of African languages to official status, the African languages still possess limited specialised terminology in some areas such as finance. This lack of such terminology and resources to deal with it remains a complex problem. The study put forward the proposition of using parallel

corpora from translation studies to create specialised bilingual dictionaries so as to develop and standardise specialised terminology in African languages. This research therefore investigated the use of parallel corpora as a useful resource for the compilation of specialised dictionaries in isiXhosa. It is important to note that corpus-based bilingual LSP lexicography is advanced in European languages whilst it is limited or non-existent in some African languages. This has resulted in difficulties when accessing information for communicative reasons.

In chapter 2, previous works and recent studies on corpus-based translation studies were reviewed in order to situate the present study. This chapter was divided into two sections. The first section was an overview of research conducted in South Africa. It was found that corpus-based research was a new phenomenon that looked at the use of corpora on various linguistic enquiries. A notable number of corpora were designed to investigate the nature of translations, translator's style, translation strategies and norms. There were very few studies conducted on the use of corpus in dictionary making; these dealt with monolingual and general-purpose dictionaries. In a review of global trends, it was proven that corpus-based research and the use of corpora are at an advanced stage in European languages, particularly English, whilst they are still lagging behind in African languages. However, specialised bilingual lexicography is still in its infancy even in English, and the situation is worse in African languages like isiXhosa. Due to this situation, some African scholars proposed the acceleration of, and motivated for the compilation of specialised dictionaries in all African languages for language and terminology development.

Because this study is located in CTS, the development of translation studies from linguistic-based theories to descriptive theories and later to corpus-based theories was investigated. In the past, translation was used for language learning within linguistic principles. One finding related to the varied views on what translation is. The earliest linguistic theorists defined translation as the transfer of language from the source text to the target text. The two concepts that were dominant in the linguistic era were 'meaning and equivalence'. Amongst the different kinds of equivalence, Nida and Newmark's dynamic equivalence appears to be a better form of equivalence though both were heavily criticised by scholars. The shift of focus from source text as the centre of translation to the target text was important for the growth of translation studies. This evolution improved the translation process with a view to producing a

comprehensible translation product: a target text that considers the function and cultural situation of the target reader. Following this approach were functionalism, descriptive translation studies and the cultural studies among others. The proponents of functionalism, Nord and Vermeer vehemently rejected the notion of equivalence and came up with the 'skopos theory and translational theory' which have been most instrumental in this current study. These two functionalist theories emphasised the function of a translated text within a cultural system.

Descriptive translation studies as the theoretical framework that underpins the study was discussed. The three research areas of DTS as an independent discipline, namely, the product, function and process were explicated. This provided an insight on what is involved in translation. The dialectical nature of DTS made an influence in this study as it focused on applied translation studies whose concern are applications to the practice of translation. The link between DTS and CTS was also discussed and showed that the two discilplines are both empirical

This study was influenced by CTS as such, the researcher traced the influence of corpus linguistics on translation and lexicography. Different types of corpora were discussed. Parallel corpora became the most appropriate type for the current study, because of its composition of the original texts and their translations. The role of parallel corpora in corpus–based translation studies was expounded. This prepared for the compilation of English-isiXhosa Parallel Corpus in Chapter 3.

The second section of this chapter looked at the development of lexicography and the influence of corpus linguistics. The interdisciplinary nature of the present study justified the brief overview of lexicographic theories that explained the types of dictionaries, their organisation and the ideal dictionary entry of a bilingual dictionary. This theoretical background assisted in the selection of corpus data that is relevant for bilingual dictionaries. The concept of bilingual lexicography and what should be included in a specialised bilingual dictionary was detailed. Of significance to this study was the functional lexicographic theory which was found to be similar to the functionalist theory of translation. The functionalist theories place emphasis on the user's needs and the function of the translation or lexicographic products. The review of relevant literature in translation and lexicography identified the current debates on the links between the two disciplines whose foundation was based on linguistics. In this section a brief overview of isiXhosa lexicography, the morphological system and the orthography was

also sketched out. Although the practice of lexicography in isiXhosa is commendable, corpus lexicography is still underdeveloped. The brief overview on the orthography was the preparation for the analysis of corpus data.

In sum, this chapter was able to contextualise the study, inform the reader about relevant studies conducted locally and around the globe, theories that underpin the study and methodological justification chosen to achieve the aims of the research.

Chapter 3 focused on the research design, data collection and analysis methods. In addressing the aims of the study, public documents on financial information were collected and used to create an English-isiXhosa Parallel Corpus. The texts used included annual reports, annual financial statements, information brochures and articles on financial matters. The creation and analysis of EngXhPC was preceded by various steps. Before the English-isiXhosa parallel texts were uploaded to ParaConc, they had to be converted into a plain text format. In order to get accurate results an alignment process was undertaken. All the steps that were followed during uploading, alignment and cleaning of the corpus were detailed and illustrated by means of screenshots. ParaConc was used to perform a quantitative analysis and was very useful in retrieving translation equivalents, frequency counts and concordances. The viability of ParaConc in term extraction was tested and proved. In a qualitative analysis conducted by the researcher, other types of information such as example sentences and collocations were obtainable from a close study of concordances. The bilingual information for lexicographic purposes was then identified and depicted by means of screenshots. It was found in this chapter that ParaConc could successfully achieve the goals stated in chapter 1.

The aim of **Chapter 4** was to present and analyse corpus data that was extracted from the EngXhPC. The findings related to research questions were then discussed. ParaConc performed different functions that enabled the retrieval of different kinds of information categories. The evidence was presented and discussed under headword lists, translation equivalents, concordances and translation strategies. The role of a parallel corpus was clearly demonstrated. The semi-automatic analysis of EngXhPC did not only generate translation equivalents which comprise the obvious information, but the examination of concordances also revealed that words in context constitute the unique output of corpus query. The typical collocations and clusters of words were easily identifiable in concordances which displayed KWIC. The ability of ParaConc to

release statistical information, sorting and pinpointing frequencies, is indicative of the benefits of corpus-based bilingual lexicography that South African indigenous languages such as isiXhosa can draw from. We have seen that the type of information extracted from this parallel corpus can improve the dictionary entries of bilingual dictionaries in specialised domains. The lexicographer's decisions will be informed by the authentic data presented by the corpus analysis tool. The speed and quality of corpus data from the aligned translated texts strengthens the use of corpora. The natural equivalents and examples are a recipe for user-friendly and comprehensible bilingual dictionaries. The term formation processes used by the translators were also identified, which resulted in a secondary finding of this research.

5.4 Contribution of the study

The use of parallel corpora as the basis for bilingual dictionaries has been widely researched in English but the findings of this study proved that the exploitation of electronic translation and lexicographic corpora in indigenous languages of South Africa, and isiXhosa in particular, is possible and that the methodology is fast, cost-effective, accurate and more efficient than the traditional method that is often used in the country. The first corpus-based bilingual dictionary of English and isiXhosa was only published in July 2014. The *Oxford Bilingual Dictionary of English and isiXhosa* is aimed at learners and covers general curriculum words. It is therefore classified as a general-purpose dictionary with a school-based audience. So far no research has been conducted on the use of specialised parallel corpora as a resource for bilingual lexicography in isiXhosa. The EngXhPC was designed and analysed to extract bilingual terminology that might be useful in dictionary making. This study, therefore, has made some theoretical, methodological and practical contributions to the field of CTS.

Theoretically, this study has made a contribution to the body of knowledge in CTS and corpus-based lexicography in African Languages. It has provided findings that support the use of translated texts as authentic texts for bilingual information. The aligned translated texts showed that the translators have used the best techniques in their translation, so that a dictionary compiler/lexicographer would be exempted from translating. The translation product, which is the outcome of a well-thought out

translation process, proved to be the result of the implementation of a functional translation theory. This study has reinforced the call for the necessary dialogue between translator and lexicographer. The findings have also contributed to the interdisciplinary approach that strengthens the links between translation theories and lexicographical theories applicable in specialised lexicography. This does not only promote interdisciplinarity but enhances synergy between different fields, corpus linguistics, translation studies, lexicography and computer science or technology.

Methodological: The authentic parallel texts that constitute the EngXhPC have proven that the computer software, ParaConc, can perform contrastive analysis of the source text (ST) and the translated text (TT). The use of EngXhPC has shifted away from the traditional, manual methods of dictionary making which are entrenched in African languages. EngXhPC can be interrogated rapidly, to extract authentic information which is the product of a theory-based translation process. The findings on the use of corpus-based lexicography and the use of a parallel corpus will contribute to the increase in and improvement of specialised bilingual dictionaries that are multi-functional, to address the needs of different users. Lastly, the study can contribute to knowledge generation: it can be used by other researchers, translation and lexicography students, teachers and linguists to better understand the use of technology in resource development.

In practise, this study will contribute to the development and standardisation of finance terminology. The type of terms and translation equivalents and their frequencies help in identifying those words that are commonly used in isiXhosa. The frequent words and highly frequent spellings facilitate the process of standardisation. It was found in the study that some spelling variations were not in line with orthography and spelling rules of isiXhosa.

5.5 Limitations of the present study

This study focused on the development of an English-isiXhosa Parallel Corpus that was used to extract bilingual information that can be used to create bilingual dictionaries. The study was limited to the field of finance and the two stated languages. An in-depth study of how ParaConc, a multilingual concodancer that can be used to interrogate parallel corpus was carried out. During the data collection phase, the researcher encountered a number of challenges and these are stated below.

IsiXhosa is the official language in the Eastern Cape, Northern Cape and Western Cape. Of these, two provinces that were selected were the Eastern Cape and the Western Cape. Although the Eastern Cape was chosen, most of the identified texts did not have translated versions. It was found that in most annual reports the finance section was not translated and the researcher was obliged to collect texts that have translated versions (cf. section 3.4.4). Should this information have been available, the size of the EngXhPC would have been fairly large. Nevertheless, it is important to state that the size of the corpus does not affect the quality of the study. The methodological limitation was also noted in the corpus analysis tool. ParaConc's semi-automatic alignment was a limitation that was beyond the control of the researcher. In the alignment of parallel texts and analysis of corpus, manual work was done by the researcher and this was time-consuming. As stated in the study, translations are not foolproof as such, the parallel corpus is bound to have errors and should therefore be used with caution. The errors occur as a result of different translators with different skills and knowledge sets. Insufficient resources also affect translations. All the aforementioned facts show that the lexicographers need to clean the corpora before use.

5.6 Recommendations

The main issues that have emerged from the findings of this study: translations are the best sources of specialised terminology and resource development. The EngXhPC as stated in section/paragraph 5.4 has made an invaluable contribution in the development of isiXhosa parallel corpora of finance terms. The translation of specialist texts into African languages has improved in the past decade due to the implementation of multilingualism, but those translations are not organised nor

collected. The main finding of the current study was that parallel corpora can be exploited to extract bilingual information that can be useful in creating specialised dictionaries. This study recommends:

- **Resource development**: IsiXhosa like many African languages is lacking resources in specialised fields yet Africa is now part of the global world that is becoming more and more technical and scientific. In order to meet the communicative needs of the African societies, there is a need to develop resources such as specialised dictionaries, terminology banks, corpus-based glossaries, spellcheckers etcetera. It is therefore recommended that all translated texts should be digitised and made available to a broader community in the form of term banks. This would also improve terminology management and access. There are many terminology development endeavours in most African languages but these are not managed properly.
- **Terminology and translation technology**: It is recommended that more research be carried out in CTS and in terminology development and translation technology. More structured and massive (parallel) translation corpora should be designed in African languages. This could help in the development and use of indigenous languages in scientific fields. Parallel corpora, for example are being used in computer-aided translations; the more available these resources the better the chances of intellectualising the indigenous languages (Moropa and Shoba, 2017).
- **Interdisciplinarity and collaboration**: This study showed that interdisciplinary approach can be used to solve linguistic problems, as such the researcher recommends that more innovative and interdisciplinary approaches be used in the field of linguistics, translation and lexicography so that the different needs of users can be met. A stronger dialogue between translators, terminographers and lexicographers is affirmed. Translators and lexicographers should have mutual interests in each other's work, because there will be no sound bilingual dictionaries without solid translations.
- **Translator training:** The training of translators should be ongoing becauce the face of translation has changed due to new theories and techonological advances. Some of the translations that were used in creating EngXhPC

displayed translation problems as a result of lack of terminology, lack of resources and not researching for terms even if these were available in other reference works. The researcher recommends that isiXhosa translators should be trained so that they centralise research in their line of work.

5.7 Future research

The EngXhPC could be extended to include all finance-related genre and carry out further investigations on context-dependent phrases and multiword units that became dominant in the concordance outputs of ParaConc. Another future direction that might be taken from the present study is the exploitation of the EngXhPC in designing an IsiXhosa-English English - isiXhosa bidirectional specialised dictionary. A greater need for the development of bilingual dictionaries with headwords in isiXhosa is arising. Researchers from other African languages can utilise this methodology to create dictionaries in different specialised fields such as health, law, technology, commerce, business etcetera. This will be beneficial to many African languages.

List of sources

Aijmer, K, and Alternberg, B. 1996. Introduction. In Aijmer, K., Alternberg, B. and Johansson, M. (eds.) *Languages in contrast: Papers from a symposium on text-based cross-linguistic studies*. Lund*:* Lund University Press, pp.11-16.

Alberts, M. 2013. Developing legal terminology in African languages as aid to the court interpreter: A South African perspective*. Lexikos,* 23:29-58.

Al-Kasimi, A. M. 1971*. Linguistics and bilingual dictionaries*. Leiden: E.J. Brill.

Atkins, B. T. and Rundell, M. 2008. *The Oxford guide to practical lexicography*. Oxford and New York: Oxford University Press.

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

Baker, M. 2011. *In other words: A coursebook on translation*. London and New York: Routledge.

Baker, P. 2010. *Corpus methods in linguistics*. In Litosseliti, L. (ed). *Research methods in linguistics.* London: Continuum International Publishing Company, pp. 93-113.

Bannock, G., Manser, W. 2003. *The Penguin international dictionary of finance*. London**:** Penguin Books.

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* (draft 3/03). Houston: Rice University.

Barlow, M. 2008. *ParaConc and parallel corpora in contrastive and translation studies*. Houston: Athelstan.

Basssnet, S & Lefevere, A. 1990. *Translation, history and culture.* London: Routledge*.*

Bejoint, H. 2010. *The lexicography of English*. Oxford: Oxford University Press.

Bennet, G.R. 2010. *Using corpora in the language learning classroom. Corpus Linguistics for Teachers.* Michigan: Michigan University.

Bergenholtz, H and Tarp, S. 1995. *Manual of specialised lexicography*. Amsterdam: John Benjamins Publishing Company.

Bergenholtz, H. and Tarp, S. 2003. Two opposing theories*:* On H.E. Wiegand's recent discovery of lexicographic functions*. Hermes, Journal of Linguistics,* 31:171-196.

Bergenholtz, H. and Nielsen, S. 2006. Subject-field components as integrated parts of LSP dictionary. *Terminology* ,12:281-303.

Bergenholtz, H. and Gouws, R. H. 2012. What is lexicography? *Lexikos,* 22: 31-42.

Biber, D. 1993. Representativeness in corpus design. *Literary and Linguistic Computing,* 8 (4):243-257*.*

Biber, D., Conrad, S. and Reppen, R. 1998. *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University.

Bowker, L.1998. Using specialized monolingual native-language corpora as a translation resource: a pilot study. *Meta*, 43(4):631-651.

Bowker, L. and Pearson, J. 2002. *Working with specialised language*: *A practical guide to using corpora.* London and New York: Routledge.

Bowker, L. 2003. Corpus-based applications for translator training: Exploring the possibilities. In Granger, S., Leroti, J. and Petch-Tyson, S. (eds.). *Corpus-based approaches to contrastive linguistics and translation studies*. Amsterdam/New York: Rodopi, pp.169-183.

Bowker, L. 2010. The contribution of corpus linguistics to the development of specialised dictionaries for learners. In Fuertes-Olivera, P.A. (ed.). *Specialised Dictionaries for Learners.* Berlin**:** Walter de Gruyter, pp.155-167.

Catford, J.C. 1965. *A linguistic theory of translation: An essay on applied linguistics.* London: Oxford University Press.

Chabata, E. 2013. The language factor in the development of Africa: A case for the compilation of specialised dictionaries in indigenous African languages. *South African Journal of African Languages,* 33(1):51-58.

Cheung, A. 2013. A history of twentieth century translation theory and its application for Bible translation. *Journal of Translation,* 9(1):1-15.

Conrad, S.M. 1999. The importance of corpus-based research for language teachers. *System,* 27(1):1-18.

Creswell, J. W. 2014. *Research design,* 4th edition. Los Angeles: SAGE Publications.

Crystal, D. 2010. *The cambridge encyclopaedia of language,* 3rd edition. Cambridge and New York: Cambridge University Press.

Department of Bantu Education. 1972. *Terminology and orthography No.3.* Pretoria: Department of Education.

De Schryver, M. and Prinsloo, D.J. 2000. Electronic corpora as basis for the Compilation of African languages dictionaries, Part 1: The macrostructure*. South African Journal of African Languages,* 20(4):291-309.

De Schryver, G.M. and Reynolds, M. (eds.). 2014. *Oxford IsiXhosa-English isiXhosa English Dictionary.* Cape Town: Oxford University Press.

De Vos, M., van der Merwe, K. and van der Mescht, C. 2015. A linguistic research programme for reading in African languages to underpin CAPS. *Journal of Language Teaching,* 48 (2):149-177. ISSN 0259-9570. http://dx.doi.org/10.4314/jlt.v48i2.7. Accessed on 16/04/2018.

Dictionary Unit of South African English. 2010. *Oxford South African Concise Dictionary. Southern Africa.* Cape Town**:** Oxford University Press.

Du Plessis, J.A. 1997. Morphology of the African languages: Imofoloji yeelwimi zaseAfrika. *Stellenbosch Communications in African Languages*, *Issue 6*. Stellenbosch: University of Stellenbosch.

Fathi, B. 2014. Experts and specialised lexicography: perspectives and needs. *Terminàlia núm*, 9:12-21.

Fernandes, L. 2006. Corpora in translation studies: Revisiting Baker's typology. *Fragmentos*, 30:87-95.

Fischer, A., Weiss, E., Mdala, E and Tshabe, S. 1985. *Oxford English Xhosa Dictionary.* Cape Town: Oxford University Press.

Flowerdew, L. 2004. The argument for using English specialized corpora to understand academic and professional language. In Connor, U. and Upton, T. A. (Eds.), *Discourse in the professions. Perspectives from corpus linguistics*. Amsterdam/Philadelphia: John Benjamins, pp. 11-33.

Fuertes-Olivera, P. A. 2009. The function theory of lexicography and electronic dictionaries: Wiktionary as a prototype of collective free multiple-language internet dictionary. In Bergenholtz, H., Nielsen, S. and Tarp, S. (eds.). *Lexicography at crossroads*. Bern: Peter Lang, pp. 99-134.

Fuertes-Olivera, P. A. 2009. Specialised lexicography for learners: Specific proposals for the construction of pedagogically-oriented printed business dictionaries. *Hermes, Journal of Language and Communications Studies,* 42:167-188.

Fuertes-Olivera, P. A. 2011. Equivalent selection in specialized e-lexicography: A case study with Spanish accounting terms. *Lexikos,* 21:95-119.

Fuertes-Olivera, P. A. and Nielsen, S. 2012. Online dictionaries for assisting translators of LSP texts: The accounting dictionaries. *International Journal of Lexicography,* 25 (2):191-215.

Fuertes-Olivera P. A. and Tarp, S. 2014. *Theory and practice of specialised online dictionaries: Lexicography versus terminography*. Berlin/Boston: De Gruyter. https://books.google.co.za/. Accessed on 25/08/2016.

Gentzler, E. 2001. *Contemporary translation theories*. Clevendon: Multilingual Matters.

Gauton, R. and De Schryver, G. 2004. Translating technical texts into Zulu with the aid of multilingual and/or parallel corpora. *Language Matters,* 35(1):148-161.

Godfrey, R. 1915. *Kafir-English dictionary*. 2nd edition. Alice: Lovedale Press.

Gómez Gonzalez-Jover, A. 2005. Specialised bilingual dictionary for translators (some considerations for a use-oriented approach). *Alicante Journal of English Studies/Revista Alicantina de Estudios Ingleses*. http://rua.es/dspace/handle/100455111. Accessed on 31/09/2015.

Gouws, R.H. 1996. Bilingual dictionaries and communicative equivalence for a multilingual society. *Lexikos,* 6:14-31.

Gouws, R. H. 2002. Equivalent relations, context and cotext in bilingual dictionaries. *Hermes*, 28:195-209.

Gouws, R. H. and Prinsloo, D.J. 2005. *Principles and practice of South African lexicography. S*tellenbosch: SUN Press.

Gouws, R and Prinsloo, D.J. 2010. Surrogate equivalents in bilingual dictionaries with specific reference to zero equivalence in dictionaries for African languages. *Tydskrif vir Geesteswetenskappe,* 50(4):502-519.

Gouws, R. H. 2012. Theoretical lexicography and the international journal of lexicography. *International Journal of Lexicography,* 25 (4): 450-464.

Gumbo, L. 2016. *Term creation: An analysis of the strategies used in some selected Shona specialised terms dictionaries*. Pretoria: University of South Africa.

Gxilishe, S., de Villiers, P. and de Villiers J. 2007. The acquisition of subject agreement in Xhosa. In Belikova, A., Meroni, L. and Umeda, M. (eds.) *Proceedings of the 2nd Conference on Generative Approaches to Language Acquisition North America (GALANA)*. Somerville, MA: Cascadilla Proceedings Project, pp. 114-123.

Hartmann, R. K. and James, G. 1998. *Dictionary of lexicography*. London: Routledge.

Hartmann, R.K. and James, G. 2001. *Dictionary of lexicography*. London: Routledge.

Hartmann, R.K. 2004. Lexicography and translation. In Sin-Wai, C. (ed.). *Translation and bilingual dictionaries*. Tübigen: Max Niemeyer Verlag, pp.7-21.

Holmes, J. S. 1972/2000. The name and nature of translation studies. In Venuti, L. (ed.). *The translation studies reader*. London and New York: Routledge, pp. 180-192.

Hunston, S. 2006. *Corpus linguistics*. Birmingham: University of Birmingham.

Jakobson, R. 1959/2000. On linguistic aspects of translation. In Venuti, L. (ed,). *The translation studies reader*. London and New York: Routledge, pp. 138-143.

Jixing, L. 2013. Translation definitions in different paradigms. *Canadian Social Sciences,* 9 (4):107-115.

Katamba, F. 2003. Bantu nominal morphology**.** In Nurse, D and Philippson, G. (eds.). *Bantu languages*. London: Routledge, pp. 103-120.

Kashgary, A.D. 2011. The paradox of translating the intranslatable: Equivalence in translating from Arabic to English. *Journal of King Saud University-Languages and Translation* 23:47-57. http:/www.academia.edu/988525/. Accessed on 20/10/2017.

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

Kerlinger, F.N. 1986. *Foundations of behavioural research*, 3rd Edition, Holt Rhinehart & Winston.

Kilgarriff, A. and Rundell, M. 2006. Efficient corpus development for lexicography: building the New Corpus for Ireland. *Lang Res Eval*, 40: 127-152. D01 10.1007/s10579-006-9011-7. Accessed on 8/09/2015.

Kilgarriff, A. and Kosem, I. 2012. Corpus tools for lexicographers. In Granger, S. and Paquot, M. (eds.). *Electronic lexicography*. Oxford: Oxford University Press, pp. 31-55.

Klein, J. 2010. Can the new African language dictionaries empower the African language speakers of South Africa or are they just a half-hearted implementation of language policies? http://www.euralex.org/elx_Euralex 2010/140. Accessed on 18/06/2014.

Koller, W. 1979/1989. Equivalence in translation theory. In A. Chesterman (ed*.). Readings in translation theory*. Helsinki: Finn Lectura. pp. 99-104.

Krishnamurthy, R. 2006. *Corpus lexicography*. Birmingham: Ashton University.

Kropf, A. 1899. *Kafir-English dictionary*. Alice: Lovedale Press.

Kruger, A. 2002. Corpus-based translation research: Its development and implications for general, literary and Bible translation. *Acta Theologica Supplementum*, 2:70-105.

Landau, S.L. 1984. Dictionaries: *The art and craft of lexicography*. Cambridge: Cambridge University Press.

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

Laviosa, S. 2004. Corpus-based translation studies: Where does it come from? Where is it going? *Language Matters,* 35 (1):148-161.

Law, J. 2010. *Oxford dictionary of accounting*. New York. Oxford University Press.

Law, J. 2014. *Oxford dictionary of finance and banking*. New York: Oxford University Press.

Leedy, P. D. (1997). *Practical research: Planning and design* (6th Edition). New Jersey: Prentice-Hall.

Lew, R. 2013. Dictionaries and Technology. In Chapelle, C. (ed.). *The encyclopaedia of applied linguistics.* Oxford: Wiley-Blackwell, pp.1-8.

Louw, P. 2006. Inclusion strategy for multiword units. *Lexikos,* 16: 95-103.

Louwrens, L.J. & Poulos, G. 2006. The status of the word in selected conventional writing systems: The case of disjunctive writing, *Southern African Linguistics and Applied Language Studies,* 24(3), 389–401.

Madiba, M. R. 2000. *Strategies in the modernisation of Venda.* Unpublished doctoral thesis. Pretoria: University of South Africa.

Madiba, M. 2004. Parallel corpora as tools for developing the indigenous languages of South Africa, with special reference to Venda. *Language Matters,* 35 (1):133-147.

Mafela, M. J. 2005. Meaning discrimination in bilingual Venda dictionaries. *Lexikos,* 15:276-285.

Masubelele, R. 2004. A corpus-based appraisal of shifts in language use and translation policies in two Zulu translations of the Book of Matthew*. Language Matters,* 35(1):201-213.

Melby, A.K. 2012. Terminology in the age of multilingual corpora. *The Journal of Specialised Translation*. www.ttt.org/tbx/art_melby-JoSTransNum18*.* Accessed on 20/03/2015.

McEnery, T and Wilson, A. 1996. *Corpus Linguistics*. Edinburgh: Edinburgh University Press.

McEnery, A., Xiao, M. and Tono, Y. 2006. *Corpus-based language studies*: *An advanced resource book.* London/New York: Routledge.

McEnery, T. and Xiao, R. 2008. Parallel and comparable corpora: What is happening? In Anderman, G. and Rogers, M. (eds). *Incorporating corpora: The linguist and the translator*. Clevedon: Multilingual Matters, pp. 18-31.

McEnery, T. and Xiao, R. 2016. Corpus-based study of Chinese. In Chan Sin-wai (ed.) *The Routledge encyclopedia of Chinese language*. London and New York: Routledge, pp. 438-450.

Mtintsilana, P.N. and Morris, R. 1988. Terminography in African languages in South Africa. *South African Journal of African Languages,* 8(4):109-113.

Marza, N.E. 2009. *The specialised lexicographical approach: A step further in dictionary-making*. Bern: Peter Lang.

Mini, B., Tshabe, S.L., Shoba, F.M. & Van der Westhuizen, P.N. (eds.) 2003. *The Greater Dictionary of isiXhosa (*Vol. 2, K-P). Alice: University of Fort Hare.

Mojela, V. M. 2010. Borrowing and loan words: The lemmatizing of newly acquired lexical items in Seotho sa Leboa. *Lexikos,* 20:700-707.

Mongwe, M. 2006. *The role of the South African National Lexicography Units in the planning and compilation of mulltifunctional bilingual dictionaries*. Unpublished M.A. Thesis. Stellenbosch: University of Stellenbosch.

Moropa, K. and Kruger, A. 2000. Mistranslations of culture-specific terms in Kropf's Kafir-English dictionary*. South African Journal of African Languages,* 20(1):70-79.

Moropa, K. 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, K. 2005. *An investigation of translation universals in a parallel corpus of English-Xhosa texts.* Unpublished doctoral thesis. Pretoria: University of South Africa.

Moropa, K. 2007. Analysing the English-Xhosa parallel of technical texts with ParanConc: a case study of term formation processes. *Southern African Linguistics and Applied Language Studies,* 25(2):183-205.

Moropa, K. 2011. A link between simplification and explicitation in English-Xhosa parallel texts: Do morphological complexities of isiXhosa have an influence? In Kruger, A, Wallmach, K. and Munday, J. (eds). *Corpus-based translation studies: Research and applications.* London and New York: Bloomsbury, pp. 259-281.

Moropa, K. and Shoba, F.M. 2017. *Language and terminology development in isiXhosa: A history.* In Kaschula, R.H., Maseko, P., and Wolff, E. (eds.). *Multingualism and intercultural communication: A South African perspective*. Johannesburg: Wits Unversity Press, pp. 76-91.

Mtuze, P. 1992. A critical survey of Xhosa lexicography 1772-1989. *Lexikos,* 2:165-177.

Munday, J. 2012*. Introducing translation studies: Theories and applications,* 3rd Edition. London and New York: Routledge.

Naudé, J. 2004. Representation of poetry in Afrikaans Bible translation: A corpus-based analysis. *Language Matters,* 35(1):226-247.

Naudé, J. 2005. The shaping of cultural knowledge in South African translation. *Stellenbosch Papers in Linguistics,* 36:35-36.

Ndhlovu, K. 2012. *An investigation of strategies used by Ndebele translators in Zimbabwe in translating HIV/AIDS text: A corpus-based approach*. Unpublished doctoral study. Alice: University of Fort Hare.

Ndhlovu, K. 2016. Using ParaConc to extract bilingual terminology from parallel corpora: A case of English and Ndebele. *Literator, 37(2):* https:/literator.org.za/index.php/literator/article/view/1278/2106. Accessed on 14/02/2017.

Ndlovu, M. V. 2009. *The accessibility of translated Zulu health texts: An investigation of translation strategies*. Unpublished doctoral thesis. Pretoria: University of South Africa.

Newmark, P. 1981. *Approaches to translation*. Oxford: Pergamon Press.

Nida, E. A. 1964. *Theory and practice of translation*. Leiden: EJ Brill.

Nielsen, S. 2014. Example sentences in bilingual specialised dictionaries assisting communication in a foreign language*. Lexikos*, 24:198-213.

Nkomo, D. 2008. *Towards a theoretical model for LSP lexicography in Ndebele with special reference to a dictionary of linguistics and literary terms*. Unpublished M Phil thesis. Stellenbosch: University of Stellenbosch.

Nkomo, D. 2010. Affirming a role for specialised dictionaries in indigenous languages. *Lexikos,* 20:371-389.

Nkomo, D. and Wababa, Z. 2013. IsiXhosa lexicography: Past, present and future. *Lexikos*, 23:348-370.

Nokele, A.B.B. 2015. *Translating conceptual metaphor in Mandela's* Long Walk to Freedom*: A cross-cultural comparison.* Unpublished doctoral thesis. Pretoria: University of South Africa.

Nord, C. 1991. *Text analysis in translation: theory, methodology, and didactic application of a model for translation-oriented text analysis*. Amsterdam-Atlanta: Rodopi.

Nord, C. 1997. *Translating as a purposeful activity: Functionalist approaches explained.* Manchester: St. Jerome.

Olohan, M. and Baker, M. 2000. Reporting that translated English: evidence for subconscious process of explicitation? *Across Languages and Cultures,* 1(2): 141-158.

Olohan, M. 2004. *Introducing corpora in translation stud*ies. London and New York: Routledge.

Otlogetswe, T.J. 2007. *Corpus design for Setswana lexicography*. Unpublished PhD thesis. Pretoria: University of Pretoria.

Pahl, H.W., Ntusi, D. M., and Burns Ncamashe, S.M. 1971. *IsiXhosa sebanga lematriki.* King Williamstown: Tanda Press.

Pahl, H.W., Pienaar, A.M. and Ndungane, A.T. 1989. *The greater dictionary of Xhosa* (Vol. 3 Q-Z). Alice: University of Fort Hare.

PanSALB. 2008. *Revised isiXhosa spelling and orthography rules*. Pretoria: IsiXhosa National Language Body, Pan South African Language Board.

Pass, C., Lowes, B., Pendleton, A. and Chadwick, L. 1995. *Collins dictionary of business.* Glasgow: Harper Collins.

Pastor, G.C. 2007. Lost in specialised translation: The corpus as an inexpensive and under-exploited aid for language service providers. *Translating and the computer 29. Proccedings of the ASLIB conference.* Londres: Aslib. ISBN 0851424856, 1-18.

Prinsloo, D. 2009. A perspective on the lexicographic value of mega newspaper corpora: The case of Afrikaans in South Africa. *Lexikos,* 19:225-241.

Pritchard, B**.** 2015. On multiword lexical units and their role in maritime dictionaries. *IJEAP*, 1 (4):40-64.

Probert, T and De Vos, M. 2016. Word recognition strategies amongst isiXhosa/English bilingual learners: The interaction of orthography and language of learning and teaching. *Reading and Writing-Journal of the Reading Association of South Africa* pp. 1-10. https://rw.org.za/index.php/rw/article/view/84/231. Accessed on 17/4/2018.

Procter, P. 1995. *Cambridge international dictionary of English*. Cambridge University Press.

Reiss, K. and Vermeer, H. J. 1984. *Towards a general theory of translational action: Skopos theory explained*. Manchester: St. Jerome Publishing.

Republic of South Africa. 1996. *The Constitution of the Republic of South Africa. Pretoria:* Government Printers.

Republic of South Africa. 1996. *UmGaqo-siseko weRiphablikhi yomZantsi Afrika*. Pretoria: Government Printers.

Rigual, C.C. and Calvi, M.V. 2014. Translation and lexicography: A necessary dialogue. *MoNTI 6*:37-62).

Roberts, R. and Montgomery, C. 1996. The use of corpora in bilingual lexicography. *Euralex '96 Proceedings*, 457- 464.

Rundell, M. 2012. The road to automated lexicography: an editor's viewpoint. In Granger, S and Paquot, M. (eds.). *Electronic lexicography*. Oxford: Oxford University press, pp.15-30.

Saule, B. and Aisulu, N. 2014. Problems of translation theory and practice: Original and translated text equivalence. *Procedia Social Behavioral Sciences,* 136: 119-123. http:/sciencedirect.com./science/article/. Accessed on 15/02/2018.

Sager, J.C. 1990. *Practical course in terminology processing*. Amsterdam/Philadelphia: John Benjamins Publishing.

Satyo, S. C. 1983. *Igrama noncwadi lwesiXhosa*. Pretoria: Via Afrika.

Schäffner, C. and Wiesemann, U. 2001. *Annotated texts for translation: English-German: Functionalist approaches illustrated*. Clevedon: Multilingual Matters Ltd.

Scott, M. 1999. *WordSmith tools*. Oxford: Oxford University Press.

Shoba, F.M. 2001. *The representation of phonetic-phonological information in Nguni dictionaries.* Unpublishers Master's degree dissertation. Stellebosch: University of Stellenbosch.

Sineke, T. 2005. *A comparative study of term creation process in isiXhosa and isiZulu translation of the South African Constitution*. Unpublished doctoral study. Johannesburg: University of the Witwatersrand.

Smit, M. 1996. The systematic development of Wiegand's metalexicography as demonstrated in Kleine Schriften. *Lexikos*, 12:290-310.

Snell-Hornby, M. 2006. *The turn of translation studies: New paradigm or shifting viewpoints?* Amsterdam/Philadelphia: John Benjamins.

Stark, M. 2011. *Bilingual thematic dictionaries*. Berlin/Boston: Walter de Gruyter.

Stolze, R. 2009. Dealing with cultural elements in technical texts for translation. *The Journal of Specialised Translation,* 11:124-142.

Svensén, B. 1993. *Principles and methods of dictionary-making*. New York: Oxford University Press.

Szal, M. 2014. Translating cultural elements in specialised texts. *Seria Filologiczna Studia Anglica Resoviensia* 11, 72-80. www.ur.edu.pl/file/72494/07. Accessed on 15/02/2018.

Tarp, S. 2000. Theoretical challenges to practical specialised lexicography. *Lexikos,* 10:189-208.

Tarp, S. 2008. *Lexicography in the borderland between knowledge and non-knowledge: General lexicographical theory with particular focus on learner's lexicography*. Tűbingen: Max Niemeyer.

Teurbet, W. 2001. Corpus linguistics and lexicography. *International Journal of Corpus Linguistics,* 6 (special issue): 125-153.

Teurbet, W. 2002. The role of parallel corpora in translation and multilingual lexicography. In Alternberg, B. and Granger, S. (Eds.). *Lexis in contrast: Corpus-based approaches,* Amsterdam/Philadelphia: John Benjamins, pp. 189-214.

Teurbet, W. 2007. Corpus linguistics and lexicography. In Teubert, W. (ed.). *Text corpora and multilingual lexicography*. Amsterdam/Philadelphia: John Benjamins, pp.109-133.

Tognini-Boneli, E. 2001. *Corpus linguistics at work*. Amsterdam/Philadelphia: John Benjamins.

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

Tshabe, S.L. and Shoba, F.M. 2006. *The greater dictionary of isiXhosa* (Vol.1 A-J). Alice. University of Fort Hare.

Tshabe, Nokele, A.B.B., Guzana, Z. 2008. *Isichazi-magama sesiXhosa.* Pietermaritzburg: Nutrend Publishers.

Tymoczko, M. 1998. Computerised corpora and translation studies. *Meta,* 43 (4):652-659.

Vermeer, H.J. 2000. Skopos and commission in translational action. In Venuti, L. (ed.). *The translation studies reader*. 2nd edition. New York and London: Routledge.

Wababa, Z. 2014. *Isichazi-magama seNzululwazi neMathematika.* Johannesburg: South African Heritage Publishers.

Wallmach, K. 2004. Pressure players or choke artists? How do Zulu simultaneous interpreters handle the pressure of interpreting in a legislative context*? Language Matters,* 35(1):179-200.

Wiegand, H.E. 1984. On the structure and contents of a general theory of lexicography. In: Hartmann, R.R.K. (ed*.). LEX'eter' 83 Proceedings: Papers from the International Confence on Lexicography at Exeter,* 9-12 September 1984. Tűbingen: Max Niemeyer. pp.13-30.

Zanettin, F. 2012. *Translation-driven corpora: Corpus resources for descriptive and applied translation studies.* Manchester: St. Jerome.

Zanettin, F. 2013. Corpus methods for descriptive translation studies. *Procedia – Social and Behavioural Sciences*, 95:20-32.

Zgusta, L. 1971. *Manual of Lexicography*. The Hague: Mouton.

Websites

Bona: https://www.bona.co.za. Accessed on 16/05/2017.

Census 2011. www.statssa.gov.za. Accessed on 15/12/2017

EasyEquities: https://support.easyequities.co.za. Accessed on 11/08/2017.

FSB (Financial Services Board): https://www.fsb.co.za, www.mylifemymoney.co.za. Accessed on 18/05/2017.

Merriam Webster's Dictionary online: https://www.merriam-webster.com/. Accessed on 20/09/2015.

PanSALB (Pan South African Language Board): https://www.pansalb.org/. Accessed on 27/01/2016.

ParaConc: http://www.athel.com/parapap.html. Accessed on 15/11/2016.

South African Government: https://www.gov.za. Accessed on 24/10/2017.

Western Cape Government: https://www.westercape.gov.za. Accessed on 20/12/2016

Wikipedia: https://en.wikipedia.org. Accessed on 1/05/2018.

Appendices

Appendix 1: Ethics Approval Letter

UNISA university of south africa

DEPARTMENT OF LINGUISTICS AND MODERN LANGUAGES:
RESEARCH ETHICS REVIEW COMMITTEE

31 August 2015

Ref #: TS_FMS08_2015
Mrs FM Shoba
Student #: 06915892

Dear Mrs Shoba

Decision: Ethics Approval

Name: Mrs F Shoba
Department of Linguistics and Modern Languages
PO Box 392, Unisa
Pretoria 0003
012 429 8546 / 083 577 0317

Supervisor: Dr K Ndhlovu

Proposal: Exploring the use of parallel corpora in the compilation of specialised dictionaries of technical terms: A case study of English and isiXhosa

Qualification: D Litt et Phil

Thank you for the application for research ethics clearance received on *27 August 2015* by the Department of Linguistics and Modern Languages Research Ethics Review Committee (RERC) for the above-mentioned research. Final approval is granted for the research undertaken for the duration of your doctoral studies.

For full approval: *The application was reviewed in compliance with the Unisa Policy on Research Ethics by the Department of Linguistics and Modern Languages Research Ethics Review Committee on 31 August 2015.*

The proposed research may now commence with the proviso that:

University of South Africa
Preller Street, Muckleneuk Ridge, City of Tshwane
PO Box 392 UNISA 0003 South Africa
Telephone: +27 12 429 3111 Facsimile: +27 12 429 4150
www.unisa.ac.za

1) *The researcher will ensure that the research project adheres to the values and principles expressed in the UNISA Policy on Research Ethics.*
2) *Any adverse circumstance arising in the undertaking of the research project that is relevant to the ethicality of the study, as well as changes in the methodology, should be communicated in writing to the Department of Linguistics and Modern Languages Research Ethics Review Committee Committee. An amended application could be requested if there are substantial changes from the existing proposal, especially if those changes affect any of the study-related risks for the research participants.*
3) *The researcher will ensure that the research project adheres to any applicable national legislation, professional codes of conduct, institutional guidelines and scientific standards relevant to the specific field of study.*

Note:

The reference number (top right corner of this communiqué) should be clearly indicated on all forms of communication (e.g. Webmail, e-mail messages, letters) with the intended research participants, as well as with the Department of Linguistics and Modern Languages RERC.

On behalf of the departmental RERC, we wish you everything of the best with your research study. May it be a stimulating journey!

Kind regards,

EJ Pretorius

Prof EJ Pretorius

Chair: Department of Linguistics and Modern Languages RERC
Tel: (012) 429 6028
pretoej@unisa.ac.za

University of South Africa
Preller Street, Muckleneuk Ridge, City of Tshwane
PO Box 392 UNISA 0003 South Africa
Telephone: +27 12 429 3111 Facsimile: +27 12 429 4150
www.unisa.ac.za

Appendix 2a: Request Letter to the Financial Services Board

University of South Africa
Department of Linguistics and Modern Languages
P. O. Box 392
Preller Street and Muckleneuk Ridge
Pretoria
0003

10 April 2017

The Senior Language Specialist
Language Business Unit
Financial Services Board
41 Matroosberg Road
Menlo Park, Pretoria

Dear Dr Lutamo Ramuedzisi

RE: REQUEST FOR OFFICIAL DOCUMENTS FOR RESEARCH PURPOSES

My name is Feziwe Shoba. I am a PhD student at the University of South Africa in the Department of Linguistics and Modern Languages. The title of my dissertation is "Exploring the use of parallel corpora in the compilation of specialised bilingual dictionaries of technical terms: A case study of English and isiXhosa". This study explores how parallel corpora can be used as resources to compile specialised bilingual dictionaries. I am specifically focusing on financial terms between English and isiXhosa. I, therefore seek permission to use parallel texts translated from English to isiXhosa. My focus is on recently published documents.

After collecting these parallel texts (English and isiXhosa) from your organisation, a specialised parallel corpus of financial terms in English-isiXhosa will be designed. The English-isiXhosa Parallel Corpus will be composed of diverse text types from different institutions/organisations. The purpose of this research is not to compile a bilingual dictionary but to extract bilingual terminology that can be useful in the compilation of specialised dictionaries. The collected information will only be used for academic purposes and the results can be provided at your request.

I hope to hear from you soon.

Yours faithfully

Mrs Feziwe Shoba

Appendix 2b: Request to The Western Cape Department of Cultural affairs and sport

University of South Africa
Department of Linguistics and Modern Languages
P. O. Box 392
Preller Street and Muckleneuk Ridge
Pretoria
0003

28 November 2016

Director: Arts, Culture and Language Services
Western Cape Department of Cultural Affairs and Sport
Private Bag X9067
Cape Town
8000

Dear Ms J. S. Moleleki

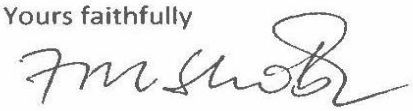
RE: REQUEST TO USE OFFICIAL DOCUMENTS FOR RESEARCH PURPOSES

My name is Feziwe Shoba. I am a PhD student at the University of South Africa in the Department of Linguistics and Modern Languages. The title of my dissertation is "Exploring the use of parallel corpora in the compilation of specialised bilingual dictionaries of technical terms: A case study of English and isiXhosa". This study explores how parallel corpora can be used as resources to compile specialised bilingual dictionaries. I am specifically focusing on financial terms between English and isiXhosa. I, therefore seek permission to use Annual Reports compiled and translated between 2010 and 2015 which are posted on your official website.

After collecting these parallel texts from your website, a specialised parallel corpus of financial terms in English-isiXhosa will be designed. The purpose of this research is not to compile a bilingual dictionary but to extract bilingual terminology that can be useful in the compilation of specialised dictionaries. The collected information will only be used for academic purposes and the results can be provided at your request.

I hope to hear from you soon.

Yours faithfully



Mrs Feziwe Shoba

Appendix 3a: Permission Letter from Western Cape Department of Cultural Affairs and Sport

From: Brent Walters [mailto:Brent.Walters@westerncape.gov.za]

Sent: 20 December 2016 02:02 PM

To: Shoba, Feziwe

Cc: Brenda Rutgers; Jane Moleleki; Deon Burger

Subject: RE: Letter of request and permission to use data

Dear Mrs Shoba

Permission for your request as detailed in your letter to us is hereby granted.

Kindly copy us on the outcomes.

Good Luck with your PhD studies.

Kind regards

Brent Walters
Head of Department
Department of Cultural Affairs and Sport
Western Cape Government

7th floor, Protea Assurance Building, Greenmarket Square, Cape Town, 8000

Tel: (021)4839501
Fax: (021)4839504
E-mail: Brent.Walters@westerncape.gov.za
Website: www.westerncape.gov.za

Appendix 3b: Approval to utilise Annual Reports of the Western Cape Provincial Treasury

Western Cape Government
Provincial Treasury

Ean Steenkamp-Cairns
Head of Communication
E-mail: ean.steenkamp@westerncape.gov.za
Tel: +27 21 483 4237 Fax: +27 21 483 5769

Enquiries: Ean Steenkamp-Cairns

7 Wale Street
Cape Town
8000

8 December 2016

Ms Feziwe Shoba
Department of Linguistics and Modern Languages
PO Box 392
Pretoria
0003

Dear Ms Shoba

APPROVAL TO UTILISE ANNUAL REPORTS OF THE PROVINCIAL TREASURY

We acknowledge your request to utilise the Annual Reports of the Western Cape Provincial Treasury for the period 2010 to 2015. These documents are created in the interest of the public and permission is therefore granted.

In the Western Cape we publish regulatory documents in English, Afrikaans and isiXhosa. Through our Language Forum, a terminology database was created that should assist you further in your research. We gladly share this terminology database with you as well.

Attached to this letter, please see the isiXhosa documents as requested.

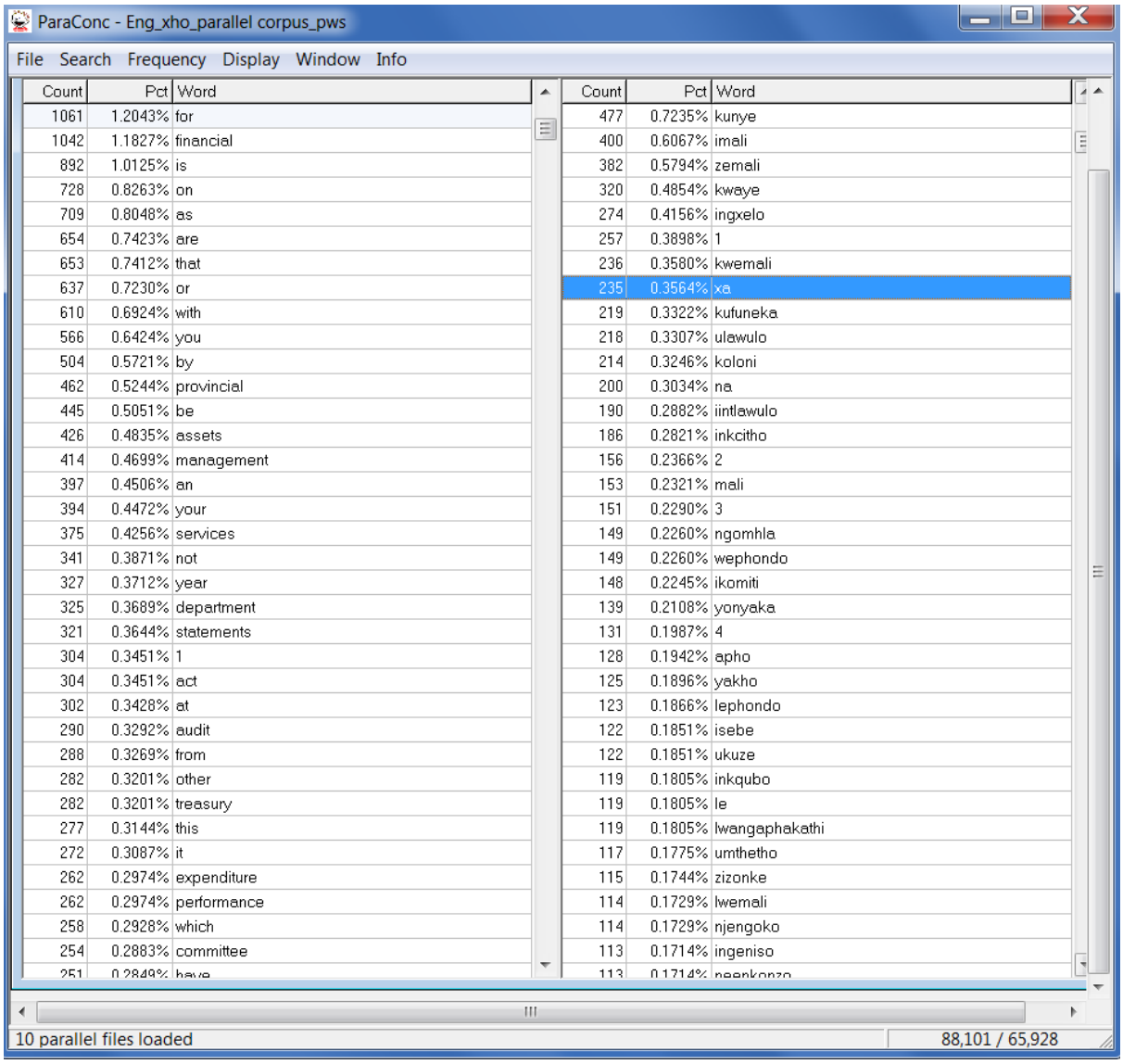
Please share the findings from your research with our department as we believe that there should be valuable learnings we can share with our Language Forum.

We wish you well with your research and the preparation of your dissertation.

Yours sincerely

EAN STEENKAMP

Appendix 4: Frequency Lists



Count	Pct	Word
1061	1.2043%	for
1042	1.1827%	financial
892	1.0125%	is
728	0.8263%	on
709	0.8048%	as
654	0.7423%	are
653	0.7412%	that
637	0.7230%	or
610	0.6924%	with
566	0.6424%	you
504	0.5721%	by
462	0.5244%	provincial
445	0.5051%	be
426	0.4835%	assets
414	0.4699%	management
397	0.4506%	an
394	0.4472%	your
375	0.4256%	services
341	0.3871%	not
327	0.3712%	year
325	0.3689%	department
321	0.3644%	statements
304	0.3451%	1
304	0.3451%	act
302	0.3428%	at
290	0.3292%	audit
288	0.3269%	from
282	0.3201%	other
282	0.3201%	treasury
277	0.3144%	this
272	0.3087%	it
262	0.2974%	expenditure
262	0.2974%	performance
258	0.2928%	which
254	0.2883%	committee
251	0.2849%	have

Count	Pct	Word
477	0.7235%	kunye
400	0.6067%	imali
382	0.5794%	zemali
320	0.4854%	kwaye
274	0.4156%	ingxelo
257	0.3898%	1
236	0.3580%	kwemali
235	0.3564%	xa
219	0.3322%	kufuneka
218	0.3307%	ulawulo
214	0.3246%	koloni
200	0.3034%	na
190	0.2882%	iintlawulo
186	0.2821%	inkcitho
156	0.2366%	2
153	0.2321%	mali
151	0.2290%	3
149	0.2260%	ngomhla
149	0.2260%	wephondo
148	0.2245%	ikomiti
139	0.2108%	yonyaka
131	0.1987%	4
128	0.1942%	apho
125	0.1896%	yakho
123	0.1866%	lephondo
122	0.1851%	isebe
122	0.1851%	ukuze
119	0.1805%	inkqubo
119	0.1805%	le
119	0.1805%	lwangaphakathi
117	0.1775%	umthetho
115	0.1744%	zizonke
114	0.1729%	lwemali
114	0.1729%	njengoko
113	0.1714%	ingeniso
113	0.1714%	neenkonzo