Chapter 5

Continuous effort, not strength or intelligence is the key to unlocking our potential.

Liane Cardes

5 Problem identification and analysis

5.1 Introduction

This chapter will focus on the possible reasons for the relatively poor retrieval performance of the empirical studies, as described in Chapter 4. In this chapter, it will be argued that two factors limit the performance of the n-gram approach where the Zulu search strings were matched to the inverted index.

The first factor is about translation ambiguity. Countless words do not have a unique translation, and sometimes the alternate translations have very different meanings. By applying every possible translation, the set of possible meanings can be greatly expanded, because some of the translations are likely to introduce additional homonyms or polysemous word senses into the second language. For instance, when untrained users enter such short queries (sometimes only one word) that it is not even humanly possible to determine the intended meaning (and the proper query translation) from the available context.

The second limiting factor regarding retrieval performance is the lack of essential dictionary terms for the correct interpretation of a query. This may occur either because the query is about a technical topic outside the scope of the dictionary, or because the user has entered some form of abbreviation or slang that is not included in the dictionary. As dictionaries (specifically designed for query translation) are developed, the effect of this limitation may be reduced. Nevertheless, it is unlikely to be eliminated, because language use is creative and new terms continuously enter the lexicon.

5.2 An analysis of errors

A detailed explanation of the process applied in this particular study was provided in Chapter 4 and only relevant sections will again be discussed for the purpose of contextualisation.

Five CLEF topics were used as a test bed, where the individual Zulu words (in the CLEF topics) were matched against the words in the monolingual electronic Zulu word list through:

- digrams,
- trigrams,
- classified s-grams,
- edit distance, and
- LCS.

All of these are approximate string matching techniques. For each of these source words, the six approximate best matches were listed for each of the five procedures. It was then manually established which one of these six words was the correct match for the source word (Section 4.4).

As saturation usually occurs around the first three (of the six) translated matches of the five different methods for the five queries (as above described), the correct Zulu word should be identified within a set of three words about 80% of the time (Section 4.4). Each word in the Zulu translations of the query topics were matched to the base-forms in the electronic monolingual Zulu word list. Based on the CLEF topics analysed, as above described, it was decided that the three best matches of the skipgram technique should be used (which identifies the correct base form in 78% of all cases).

Through a manual analysis of the matching of the queries to the original database entries, and a comparison of the official translations to the mother tongue translations two types of problems were identified (figure 5.1). They are **dictionary problems** and **translation problems**. In Figure 5.1 the two types of problems are compared in terms of the number of errors occurring with each. To categorise the errors causing the relatively poor retrieval performance, 35 queries were manually analysed. By literally counting the errors on a word-by-word basis, 169 occurrences of translation problems (in eleven sub–categories) were found, compared to 89 dictionary-related problems (in three sub-categories). This amounts to 258 instances (in 398 words matched to the index word list) of either dictionary- (34%) or translations problems (66%) in these 35 queries.

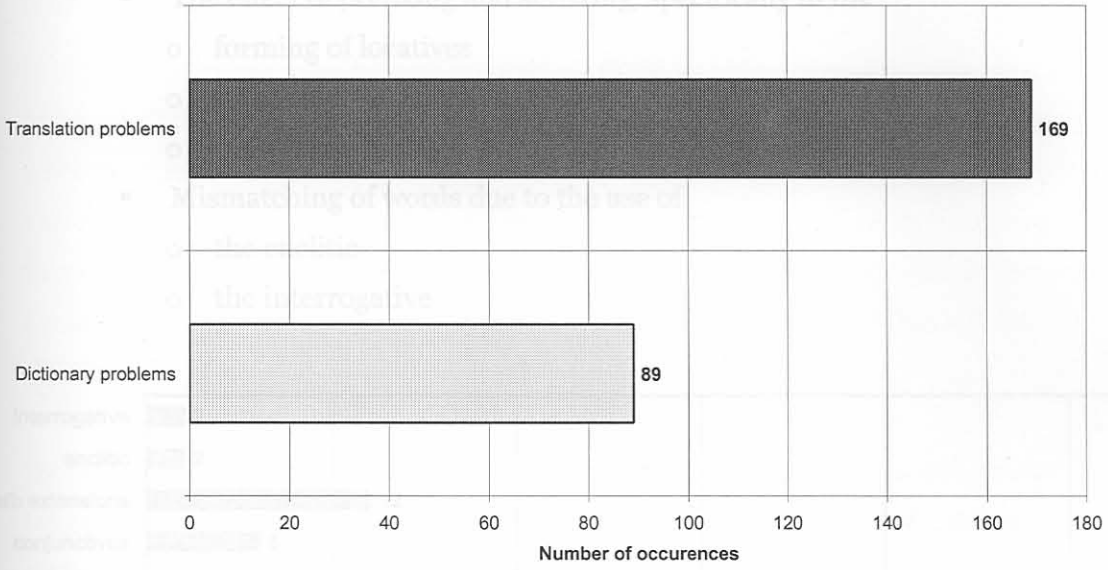

Figure 5.1 *A comparison between the dictionary problems and translation problems*

Typically, dictionaries provide several translations for a single source language word. Therefore the number of mistranslated keywords (i.e. the keywords that have incorrect meanings in the context of the topic in a CLIR query (the final translated query)) is usually high. Through the manual analysis of the queries, several **dictionary** problems were identified. These can be divided into:

- The orthographical rules of Zulu grammar;
- The frequent manner in which words are borrowed from English and Afrikaans;
- Zululised words; and
- The (mis)matching of proper names.

Apart from dictionary problems, numerous problems were also experienced during **translation**. These problems can be divided into the following:

- Paraphrasing, phrase translations and the effect of compounding;
- Changes to the noun and verb stems through word inflection, specifically
  - palatalisation
  - pre-nasalisation
  - the coalescence of vowels
  - vowel elision
- Homonyms and resulting mistranslations;

- The effect of prefixing and suffixing, specifically in the
  - forming of locatives
  - forming of conjunctives
  - verbal extensions
- Mismatching of words due to the use of
  - the enclitic
  - the interrogative

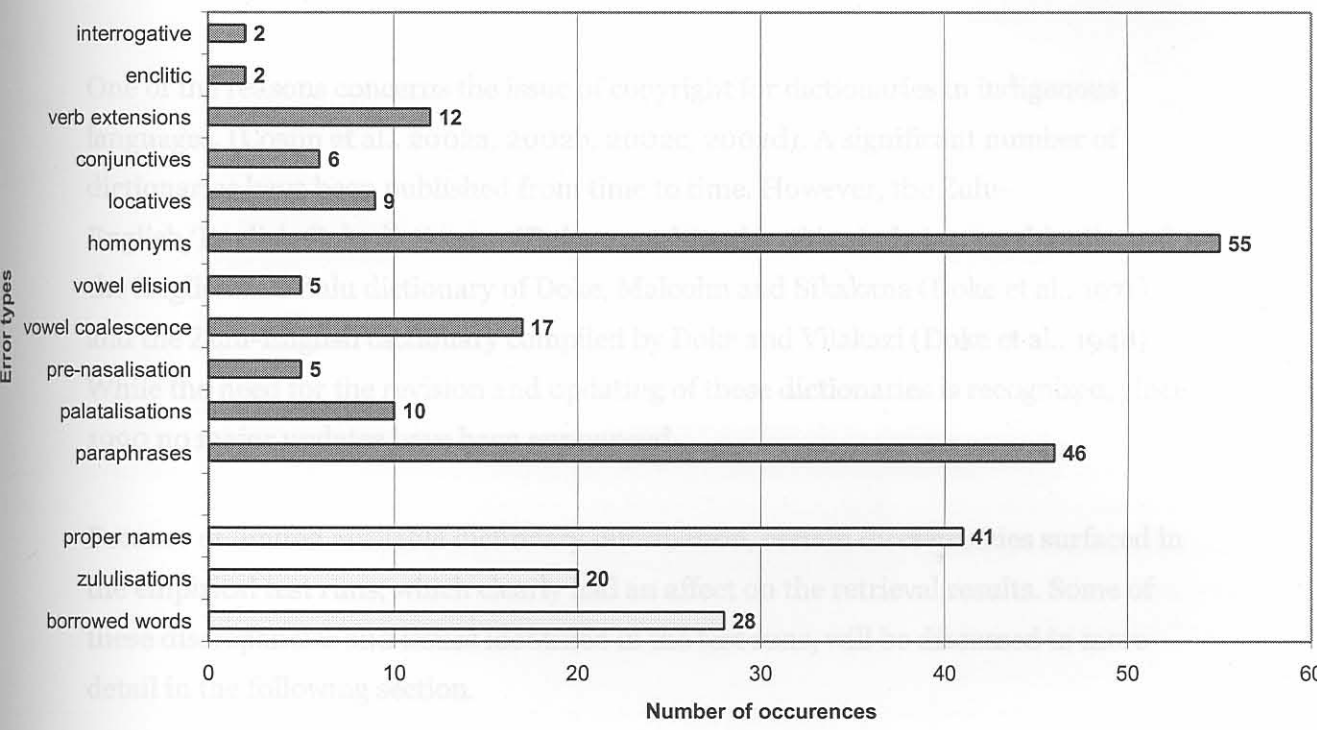

Figure 5.2 *A comparison of the types of errors encountered in the translation process*

Figure 5.2 compares the different types of errors encountered in the translation process with each other. Here, 398 words were taken from 35 queries and matched to the index word list. Of the 398 words, 258 words matched to entries found in the index word list, while there were no matches for 140 of the words. Of all the words (398), 262 were analysed and divided into two categories (dictionary and translation). Each had different sub-categories. The number of occurrences for each of the identified errors is indicated in Figure 5.2.

Each of the categories and sub-categories in Figure 5.2 will be discussed in more detail by using examples from the tables in Appendix A.

5.3 *Dictionary problems*

It is possible to improve the performance of CLIR by using more than one lexical resource, but during the research conducted for this study, other numerous problems were experienced with the available resources in electronic format. The number and variety of dictionaries in the Zulu language far exceed other indigenous languages in this country. However, currently there are no subject-specific or special dictionaries available in Zulu. This emphasised some of the errors experienced that will be discussed below.

One of the reasons concerns the issue of copyright for dictionaries in indigenous languages. (Cosijn et al., 2002a, 2002b, 2002c, 2002d). A significant number of dictionaries have been published from time to time. However, the Zulu-English/English-Zulu dictionary (Doke, 1990) used in this study is a combination of the English and Zulu dictionary of Doke, Malcolm and Sikakana (Doke et al., 1971), and the Zulu-English dictionary compiled by Doke and Vilakazi (Doke et al., 1948). While the need for the revision and updating of these dictionaries is recognized, since 1990 no major updates have been announced.

Because of limited available dictionary information, certain discrepancies surfaced in the empirical test runs, which clearly had an affect on the retrieval results. Some of these discrepancies and issues identified in the test runs, will be discussed in more detail in the following section.

5.3.1 Orthography

For a long time, Zulu was an oral language, and it is only recently (since 1848) that people have begun to document it (Wilkes and Nkosi, 1995). It is also only the third African language (besides Xhosa and Setswana) to appear in print. It is important to stress why a language needs to have a literary form. The lack of a written version can be costly to a language, as the integration with modern societies forces native people to adopt a recorded method of communication (Wilkes and Nkosi, 1995).

According to Wilkes (2001, 1988), the orthography of a language specifies a number of aspects about the writing of a language. Especially in how the speech sounds of a language should be written, as well as what the recognized words in the language are so that autonomous words are written separately from one another.

In 1934, the first Zulu orthography was officially recognized. Some of the most prominent changes occurred in 1950, when the Zulu orthography was revised and further amendments added in 1955. The latest edition, *Terminology and Orthography No. 4 for Zulu*, was issued in 1993. One of the implications of the updating and revision of the orthography is that certain discrepancies exist. Therefore, the following rules (Wilkes, 2001) pertaining to the writing of the sound system of Zulu are enforced, are indicated in Table 5.1:

Table 5.1 *The different writing rules related to the sound system of Zulu*

Rule number	Rule description	Example in text	Comments
Rule 1	The bilabial implosive, previously written as *ɓ* is represented by b.	*ubulala* (Table A1.1) *ubuzulu* (Table A3.2) *ubungeno* (Table A6.2) *ubukhanda* (Table A10.1) *ubuhlungu* (Table A13.1) *ubulima* (Table A18.2)	In the dictionary (Doke et.al, 1990) used for this study, some words were still written with a *ɓ*. This made it difficult in some instances to correctly match the dictionary entries.
Rule 2	The devoiced bilabial explosive b will now be written as *bh*.	*imibhalo* (Tables A2.1, A2.2) *zebhola* (Tables A8.1, A8.2).	It does occur that mother tongue translators are not aware of changes to the writing system, and therefore still write words in the "old syntax".
Rule 3	The bilabial explosive in nasal compounds is written as *mb*.	*imbibika* (Table A34.2)	In most instances it is not easy to determine when the nasal sound influences the spelling of a word. This emphasizes the occurrences of palatalised words.
Rule 4	The voiced glottal fricative will be represented by *hh*.	*-hhalo* (Table A19.2) *-hholisaka* (Table 30.2) *-hhovini* (Table A35.2)	As with rule 2, the mother tongue translators are not aware of the changes to the orthography. However, this does not necessarily influence the resulting matches, but it should be taken into consideration.
Rule 5	The ejective alveolar affricate will be represented by *ts*.		This rule does not apply to the current study.

Rule 6	Where semivowels occur between vowels (meaning that two vowels come together without a glottal stop between them), a semivowel (*y* or *w*) may be inserted between these vowels.	*-nayiti* (Table A2.2) *-momoyi* (Table 3.2) *-shayisa* (Table A20.2)	This has a big impact on the identification of the stem used in matching the dictionary entry. It also frequently occurs in verb extensions. For more examples, see section 5.4.4.3.
Rule 7	The hyphen will be used as follows: • When a numerical is preceded by an inflected prefix. (Cardinal numbers are to be written without any concords preceding them. For example *umbuzo 10, isifundo 2*, but ordinal numbers preceded by *ngomhla ka* will be written with a hyphen. For example *ngomhla ka-10*). • To separate two vowels coming together with a glottal stop between them	*wezingcweti zebhola lomhlaba ngo–1994* (Tables A8.1, A8.2). More examples can be found in Tables A16.1, A16.2, A22.1 and A22.2. *i-UN* (Table A2.1) *i-El Nino* (Table A3.1) *i-Europe* (Table 11.1)	There is no specific rule about the use of the hyphen in a grammatical sense – it is mostly enforced for better readability in the text.
Rule 8	The apostrophe is used to indicate elision, but is not relevant to this study.		

The orthography of Zulu (including the above rules) is based on phonetic principles as with any other language. Issues regarding pronunciation, foreign acquisitions, the use of the hyphen, stress-indication and the use of capital letters will be discussed in the following sections.

5.3.2 Borrowed words

It is only recently that a language like Zulu has begun to be subjected to the westernised world. Because of this it does lack certain words that are standard in technologically advanced societies. For example, words associated with inventions like 'electricity', 'telephone', and 'computer'. Therefore, most of the borrowed words

found in the text are not found in the dictionary yet, although some of them have been included as new updates to dictionaries.

Zulu for example, has no known word for "computer virus", and because of this it has been borrowed from the English language. Even the sound of the letter 'r' has entered the Zulu language only through borrowed words such as 'radio', 'shirt' and 'computer'.

Zulu has two categories of borrowed words. Category one refers namely to words that are **adapted to Zulu pronunciation**, and category two refers to words that remain unchanged with only **class prefixes added to conform to the grammatical system**. In the first category, words are borrowed from another language, adapted to Zulu pronunciation according to the phonological system and eventually included in the dictionary. This is known as *Zululising* the foreign words (also see Section 2.3.1.4). An example of a borrowed word adapted from English, but which has not yet been included in the dictionary is *amakhemikheli* (Table A1.1). It sounds similar to the English equivalent 'chemical'. The Zulu word for chemical that do appear in the dictionary is *-phathalene* or *-thakiweyo*, and they are actually adjective forms.

Another example of a borrowed word is 'computer'. In Table A14.1 and Table A18.1 (Appendix A) the word is mismatched, because computer is Zululised to *ikhompiyutha*. Furthermore, the mother tongue translation does not have a word, so the rule for borrowed words suggests that a class prefix should be applied to the English word. This results in *'i-computer'*. Note that the class prefix is actually class five (*ili-*), but the *-i(li)* takes a silent form, as is the case with almost all borrowed words.

The same can be applied to 'virus'. The English equivalent is Zululised and a class prefix -ama is added to conform to the Zulu grammatical syntax. The result (*amavayirasi*) would not be found in the dictionary. Interestingly enough, the mother tongue translation do translate the word 'virus' into the following phrase: 'poison of the computer'.

In the following examples (Table 5.2), the Zulu words borrowed from the English language are displayed. Note, that although the spelling of Zulu words differs quite substantially from that of the English spelling, the pronunciation is very similar.

Table 5.2 Zululised borrowed words

CLEF topic	English word as in CLEF	Zulu equivalent
C042 (Table A2.1)	soldiers	*amasosha*
C051 (Table A8.1)	ball	*ibhola*
C060 (Table A15.1)	politics	*epolitiki*
C068 (Table A21.1)	synagogue	*isinagogo*
C074 (Table A25.1)	tunnel	*ithonela*
C075 (Table A26.1)	court	*ikantolo*
C075 (Table A26.1)	Germany	*iJalimane*
C078 (Table A29.1)	September	*uSepthemba*

It is interesting to note that the word 'synagogue' can also be paraphrased (see Section 5.4.1) into *nezindlu zesonto zamaJuda*, and when translated it becomes 'the dwelling place on Sundays of the Jews'.

Table 5.3 indicates the second category of borrowed words in Zulu. In this category, words are taken as is from the host language (in this instance English), and only a class prefix is added so that the word will conform to the Zulu grammatical system. However, in adding the class prefix some borrowed words are hyphenated as indicated in the table. There are no clear rules for the use of the hyphen. From Table 5.3, it can be seen that *iTour de France* does not take a hyphen, while *i-El Nino* does. This is because the hyphen in the instance of i-El Nino is used between the two vowels (see Table 5.1, Rule 8) to simplify the reading and writing of the word. The mother tongue translator even attempted to translate *Tour de France* to 'the journey that is endless'. Another example is the word 'computer'. In the instance of the official translation it is Zululised to become *ekhompiyutha*, while the mother tongue translation results in *i'computer'*. In this instance the borrowed word is put in inverted commas. This is done to simplify the reading and writing of the word, and to indicate that the word is indeed in its borrowed (English) form. There are no specific rules that justify the inconsistent use of the hyphen or inverted commas. It is the same with *igreen power* (compare with *i'green power'* or *i-green power* or *i-'green power'*) and *iMad Cow Disease* (compare *i-Mad Cow Disease* or *i-'Mad Cow*

Disease'). It may be assumed that the inverted commas are used to indicate that the class prefix is added to the phrase as a whole, and not just the first word. Grammatically, any of the written forms could be correct in terms of use.

Table 5.3 Borrowed words with a class prefix added

CLEF topic	English word as in CLEF	Zulu equivalent
C043 (Table A3.1)	El Nino	*i-El Nino*
C044 (Table A4.1)	Tour de France	*i-Tour de France*
C059 (Table A14.1)	Computer	*i-'computer'*
C086 (Table A33.1)	green power	*i-'green power'*
C088 (Table A34.1)	Mad Cow Disease	*i-'Mad Cow Disease'*

** In Zulu, borrowed phrases (common nouns) are usually indicated by using inverted commas. (See Section 4.5).

As indicated in Figure 5.2, there were 28 instances where borrowed words occurred, making it the fourth highest type of error to be encountered in the analysis of the words.

5.3.3 Proper names

Proper names in Zulu are usually found in class 1a, as indicated in section 3.3.1. However, it can be added that the singular and plural forms of the category of personal names in proper names in class 1a are indicated by an u- (singular) or o- (plural). Thus, uJames can become oJames (James and the others). This implies that all proper names will take the appropriate prefix, as indicated in Table 5.4.

Table 5.4 *Proper names in Zulu*

Category of proper name	Example of Zulu proper name
Personal names	*uJames, uSusan*
Names of the Deity	*uNkulunkulu* (God), *uJesu* (Jesus)
Books of the Bible	*uGenesise, uEksodisi, uEzra*

Days of the week	*iSonto* (Sunday), *uMsombuluko* (Monday), *uLwesibili* (Tuesday)
Titles of people	*uDokotela Nkomo, uProfesa Dube*
Geographical names (not place names)	*i-Amazon, i-Antatika*
Names of organizations, historical events, and government departments	*iNhlangano yeZizwe* (United Nations), *uMkhosi wePhasika* (Passover), *iMfuduko Enkulu* (the Great Trek), uKhisimusi (Christmas)
Place names	*KwaZulu, eGoli, eManzimtoti, eThekwini*

As indicated in Table 5.4, if the proper name starts with a vowel, a hyphen is normally added between the prefix and the proper name. This is also true for the borrowed words found in Appendix A. For instance *i-El Nino* (Table A3.1) and *u-Indurain* (Table A4.1). Furthermore, it must be added that in the experiments conducted, the proper name was not matched through approximate string matching. To illustrate this, take 'Haiti' as an example: Haiti is not matched in the dictionary, but because it is a proper name, it takes the *lase-* in front of the stem. The reason for this is that, as a rule, most place names take a *lase-* or *kwa-* as a prefix, indicating 'the place of' or 'the country of'. 'US' though, are matched to *Melika*, since the word is actually found in the dictionary. For 'UN', the system matches the word to a paraphrased form 'the place of nations that is united' or *weNhlangano yeZizwe*.

It is also interesting to note that the proper name 'France' has a dictionary entry (*eFulansi*), but the mother tongue translation applies the rule for proper names by just adding a class prefix to the English word (*eFrance*).

A "stripping technique" was used, whereby prefixes and affixes were automatically identified and removed, thereby restoring the proper name to its original form (Cosijn et al., 2002a, 2002b, 2002c, 2002d). It was then possible to match the proper name directly to the database through a table lookup procedure to see whether the root of the word was spelled correctly. These words were recognized automatically by the system through the predetermined grammar rules (as explained in Sections 3.2, 3.3 and 5.1.2).

Apart from the dictionary problems, other errors were also encountered in the translation process. This was mentioned in Section 5.2, and will be discussed now.

5.4 *Error analysis of the translation process*

While locating translations in dictionaries, the following problems were encountered: missing words, proper name translation, spelling conventions and technical term coverage.

It was quite evident in this study, as the dictionary did not contain some of the required words, and it did not include all the possible meanings of the words it referred to. However, most importantly, it was impossible to resolve translation ambiguities by just using a dictionary. The problem is that if two possible translations were given for the query term, one would not necessarily know which translation the system would choose (Cosijn et al., 2002a, 2002b, 2002c, 2002d). Also, by using only a simple dictionary look-up, a high degree of ambiguity is introduced thereby making it difficult to identify translation equivalents.

Some of the more specific errors that were identified and indicated in Figure 5.2, will be addressed now.

5.4.1 Paraphrasing

According to the Collins Dictionary of the English Language (Hanks, 1983), paraphrasing is "an expression of a statement or text in other words, especially in order to clarify". It is common to use paraphrasing in indigenous languages like Zulu, because some of the technical and scientific terminology do not exist.

In our experiments (Cosijn et al., 2002a, 2002b, 2002c, 2002d), a sworn translator translated the CLEF English topics into Zulu, and a Zulu mother tongue speaker then examined these translations. To demonstrate the affect that the fundamentally different vocabularies of English and Zulu could possibly have on dictionary translation, mother tongue speakers (who were not translators) again independently translated 35 of the CLEF topics into Zulu. The first set of translations from English to Zulu was called 'official translations', and the second set of translations was called 'mother tongue translations'. In both sets of translations, a high incidence (18%) of paraphrasing were found (see Table 5.5). This can specifically be ascribed to the fact that English and Zulu do not really have comparable vocabularies, especially for technical terms. In Section 2.3.1.4, a study by Ballesteros and Croft (1996) on English to Spanish text and Spanish to English text, revealed a considerable loss in precision when queries were translated word-by-word, due to the absence of comparable vocabularies for the two languages.

Examples of paraphrasing have been listed below in Table 5.5. Some of the paraphrasing occurs in the official translation, while others in the mother tongue translation. In some instances, the paraphrasing was different in both sets of translations.

Table 5.5 *Paraphrasing in Zulu*

CLEF Topic	English word as in CLEF database	Zulu paraphrase	Direct translation back into English
C041 (Table A1.1)	pesticides	*amakhemikheli abulala zonke izifo ezinengozi*	Chemicals that put to death all diseases that are dangerous
C041 (Table A1.1)	pesticides	*umuthi wokubulala izinambuzane*	medicine that kills insects
C043 (Table A3.1)	weather	*nesimo sezulu*	the shape of the sky
C043 (Table A3.1)	temperature	*kwizinga lokushisa*	the process of becoming hot
C043 (Table A3.1)	rainfall	*nokunetha kwezulu*	the coming of wetness to the sky
C044 (Table A4.1)	tour	*emncintiswaneni wohambo*	the competition of the journey that is endless
C044 (Table A4.1)	reactions	*ukuphatheka kwabantu*	to deal with the feelings of the people
C046 (Table A5.1)	embargo	*ukuvalelwa kokushintshisana ngempahla*	the closing of mutual exchange of goods
C058 (Table A13.1)	euthanasia	*ukufa ngaphandle kokuzwa ubuhlungu*	death that is right of the senses that are painful
C063 (Table A16.1)	earthquake	*ukuzamazama komhlaba*	the shaking of the earth
C073 (Table A24.1)	referendum	*isicelo sokuthola umqondo wabo bonke abantu*	the application to find out the opinion of everybody
C081 (Table A31.1)	airports	*ezikhumulweni zezindiza*	the place of airplanes
C088 (Table A34.1)	'Mad Cow Disease'	*isifo samatele*	the disease of the hoofs of animals
C090 (Table A35.1)	exporters	*abantu abathengisa kumazwe angaphandle*	people that trade from the country to the outside
C090 (Table A35.1)	frozen	*eqandisiwe*	that is very cold like ice

More examples of paraphrasing can be found in Tables A8.1, A12.1, A19.1, A21.1, A28.1 and A33.1 (Appendix A).

From the above table, it seems as if the dictionary translation attempts to describe each of the words in phrases of three to four words, like temperature (process of becoming hot), or air pressure (the influence of the wind) and even rainfall (the coming of wetness to the sky). When the mother tongue translators were asked why the dictionary and mother tongue translation differ so much in terms the paraphrased words, the reply was: "we do not really speak like that".

These forms of paraphrasing are the result of directly attempting to match the Zulu words to dictionary entries. These are just a few examples that demonstrate the instances where no matches were made (for example *nesimo* and *lokushisa*) and where mistranslated matches occurred. In the instance of *nesimo*, the noun stem is actually *-mo*, and therefore no match was made. This also accounts for *lokushisa*, which is found in the dictionary beneath *ukushisa*.

Although there is no real Zulu word for 'corruption', the mother tongue translator indicated that the word *ngenkohlakalo* is generally used for corruption in the sense of "events or happenings of a criminal nature".

It is clear that the dictionary translation and mother tongue translation are different from one another. In Table A33.1 (Appendix A) for instance, the mother tongue translation for 'renewable' resulted in the following paraphrased form: 'that can be made new' (*angenziwa kabusha*).

In addition, both forms of translations paraphrased 'describing' from the original query description to 'that show the meaning of' (the original translation is *echaza ngokusetshenziswa*) and 'that joins together work' (the mother tongue translation is *ochaza ukusetshenziswa*). Interestingly enough, both Zulu matches are the same, but the English matches slightly differ.

One of the cultural aspects that play a role when moving towards cultural CLIR is illustrated in the instance where the words "green power" are mistranslated in the mother tongue translation. The mother tongue translation has nothing to do with electrical power, but rather power in the sense of strength, and then it must be of a green colour as well.

In Table A34.1, the query concerns Mad Cow disease, and detecting documents related to this sickness. In the mother tongue translation, however, the term "mad cow" is translated to "the disease of the hoofs". This is mainly because Zulus are aware of the disease, and they know it comes from animals with hoofs (and not only cows)–therefore the culturally correct translation. The dictionary translation again only applies the rule for borrowed words and adds the class six prefix to obtain "i-Mad Cow disease".

These examples in itself provide a framework for in-depth research into cultural translation and the correct interpretation of texts, as a great portion of a message is lost through the self-interpretation of texts without putting it into a desired context.

Mechanically, n-grams match most of the individual words quite well to that of dictionary entries. On a conceptual leve,l however, the result is not always good. This can be ascribed to the system of grammatical rules (see Section 3.3) that are applied to the text. When these rules are enforced in the different categories, it generate all the possible dependencies that are allowed on a conceptual level. This was evident in the experiments, as several differences are found between English and Zulu (both on grammatical and conceptual level). Furthermore, in terms of paraphrasing it was found that the association between two classes did not always contribute to the user's (in this case the translator) understanding of the text. This is one of the main problems of paraphrasing–it is a *mechanical* process of exchanging words and phrases for synonyms. In effect, it discourages careful consideration of the meaning of the text itself. Consequently, the translator may interpret the text differently, and as a result make small changes in the written text. In doing this, the translator actually changes the meaning of the original text itself. As a result, the amount of paraphrased text becomes too general or unfocused, and which was not always effectively translated or matched on a conceptual level.

5.4.2 Word inflection

The degree of word inflection varies from language to language. For a very dissimilar language pair such as Zulu and English, the difference is obvious and is necessary to be addressed to ensure a good quality translation. Therefore, a statistical n-gram matching approach was well able to address this issue.

In a study by Pirkola et. al (2001), they indexed the unrecognized words (more than 119 000 forms) of their Finnish test database as digrams (see Section 4.2 for

definition) and found "that the frequency of such digrams that act as suffixes or are part of suffixes is often high in particular at suffix positions". The following illustrates this: the *in*-gram is a genitive suffix and common at all positions of strings, but it is far more common at the last position (where it occurred more than 65 000 times), i.e., genitive position than at other positions. Pirkola's findings suggest "that n-gram based stemming in which the n-grams with high frequency at suffix positions would be down weighted, would benefit retrieval" (Pirkola et al., 2001). This is also true for Zulu, where the different verb-extensions can be assigned certain weighted values to improve retrieval results.

If the source language words appear in inflected forms, they cannot be readily translated, because they do not match dictionary headwords (which are in base forms). Zulu is a highly inflectional language (30% of the errors were because of word inflection), and matching the words in the running text to the dictionary entries is often problematic. The reason for this is the changes to the syntactic function of the word, thus being inflected for the singular and plural, for future and past tenses, as well as other tenses.

5.4.2.1 Palatalisation

In Zulu, bilabial consonants may not be followed by -w-. In some instances this semi-vowel is dropped, but in most instances the bilabial consonant gives way to a corresponding prepalatal sound. The process whereby words are changed to conform to certain inflectional forms are known as palatalisation (Bosch and Taljaard, 1988).

When the noun ends in *-bu, -bo, -phu, -pho, -bhu, -bho, -mu, -mo, -mbu, -mbo, -mpu* and *-mpo*, the sounds change as follows:

Table 5.6 *The effects of palatalisation*

CLEF topic	Sound change	Zulu word	Grammatical analysis (with source word underlined)
Co65, Co76, Co86	**b>tsh**	*ukusetshenziswa*	*uku + sebenza + izwa*
Co74	**b>tsh**	*ukugcotshwa*	*uku + gcoba*
Co76	**b>tsh**	*catshangwa*	*cabanga*
Co48, Co57	**ph > sh**	*zokukhishwa*	*za + uku + khipha*
Co44	**m > ny**	*okhonyeliswa*	*o + khomela + iswa*
Co57	**m > ny**	*abafunywana*	*aba + fuma + na*

Co60, Co64	**m > ny**	*kungavunyelwe*	*ku + nga + vuma + elwa + e*
Co46	**mb > nj**	*ukuvinjelwa*	*uku + vimba + elwa*
Co79	**mb > nj**	*ngezinjongo*	*nga + izim + banga + ongo*

Other sound changes that occur (but not as frequently as those indicated in the table), are **mp > ntsh**, **th > tsh**, **n > ny**, **nt > ntsh**, and **nd > nj**.

Approximate string matching in such complex cases does not result in any meaningful match. It is suggested that the only solution for this problem would be an effective morphological parser (Cosijn et al., 2002d).

Palatalisation occurred in at least nine of the queries analysed in Appendix A (indicated in Figure 5.2), amounting to 4% of the errors encountered.

5.4.2.2 *Pre-nasalisation*

A homorganic nasal compound consists of the *-n-* of the *in-/izin-* class and a following consonant, for example *-nt-* in *into*. Some of the nasal compounds in Zulu are not that noticeable in the practical orthography of Zulu (Bosch and Taljaard, 1988). The following examples will suffice:

Table 5.7 *The effects of pre-nasalisation*

Nasal sound	Occurrence in zulu word	English meaning
ph > mp	impi	army / war
bh > mb	imbuzi	goat
f > mf	imfene	baboon
y > ny	inyama	meat
sh > ntsh	intshe	ostrich
kh > nk	inkabi	ox
g > ng	ingozi	danger
s > ns	inso	kidney
hl > nhl	inhlwathi	python

Other nasal compounds include **ph > mp**, **th > nt**, **d > nd**, **z > nz**, **hl > nhl** and **qh > nq.**

Occurrences of pre-nasalisation were found in five of the query words analysed in Appendix A. This is indicated in Figure 5.2, which amounts to 2% off all errors encountered.

5.4.2.3 Vowel coalescence

The Zulu basic vowel *a* (usually found in the instrumental *nga-* and the adverbial formative *na-*), when followed by the vowels *i* or *o*, may at times coalesce to form *nge-* or *ngo-*, as indicated in Table 5.8.

Table 5.8 *The three principal cases of vowel coalescence*

a + i = e
a + u = o
a + a = a

The inflected word forms resulting from vowel coalescence (7%) were found in more than seven of the queries in Appendix A, as indicated in Figure 5.2.

5.4.2.4 Vowel elision

Apart from the grammatical elision of initial vowels (when forming vocative interjections or after adverbial formations), vowel elision occurs when two word groups are combined to form a new word group. There are two kinds of elision:

- Elision of the initial vowel of the second word, for example *lo* + *umuntu* = *lomuntu* (this person) or *leso* + *isihlalo* = *lesosihlalo* (that chair).
- Elision of the final vowel of the first word, for example *inkosi* + *enkulu* = *inkosinkulu* (big chief) or *bonke* + *abantu* = *bonkabantu* (all people).

Other examples specific to the different noun classes are shown in Table 5.9.

Table 5.9 *The occurrence of vowel elision in the different classes*

Class	Vowel elided	Zulu example	English meaning
1st person singular	ngi + -eba > ngeba	Ngeba imali.	I steal money.
1st person plural	si + -osa > sosa	Sosa inyama.	We roast meat.
Class 2	ba + -ala > bala	Abafana bala ukudlala	The boys refuse to play.
Class 5	li + -eqa > leqa	Ibhubesi leqa isango.	The lion jumps over the gate.
Class 7	si + -opha > sopha	Isilonda sopha kabi.	The wound is bleeding badly.
Class 8	zi + -ephula > zephula	Izigebengu zephula amfasitele.	The gangsters break the windows.
Class 10	zi + -onga > zonga	Izintombi zonga imali.	The girls save money.

Examples of vowel elision can be found in Tables A4.1, A12.2 and A13.2. The occurrences of vowel elision (2%) compared to other error types analysed are shown in Figure 5.2.

5.4.3 Homonyms

Homonyms are words that have the same form or sound, but the words have different meanings. In Zulu, the difference in meaning is sometimes only conveyed by tone (Canonici, 1995). Tone, described by Doke (1968) as "the sequence of musical pitch upon the syllables of words" is a very important element in Zulu. This is because tone in Zulu is semantic–it is a deciding factor in the meanings of words. It is difficult for a machine to recognize where tone is indicated in a word, since tone is usually created by means of an acute symbol.

The Zulu language has a relatively high occurrence of homonyms (more than 50 occurrences in over 90% of the queries analysed). Compared to the other problems experienced in the test runs, homonyms amount to 21% of all problems, which make this the most frequent error encountered in the analysis. Table 5.10 indicates some examples of homonyms found in the analysed queries, with up to three different meanings produced for a word, depending on where the tone is. This indication of tone would, however, not be detected by the machine during the translation process.

Table 5.10 Examples of homonyms

CLEF topic	Zulu word	Dictionary meaning (English)	Second meaning	Third meaning
C043	*-netha*	get wet, be soaked, leak	be sluggish, inactive, drowsy	
C043	*-zulu*	sky, lightning, weather	Zulu language and mannerisms, member of Zulu nation	intoxicating drink made from the juice from the heart of the ilala-palm
C043	*-hlaba*	earth, world, land	something of no value	
C043	*-phumela*	Cluytia Pulchella, a small scrub of the Euphorbia-family, used as a love-charm	after-effect	come out, rise for, leave for, discharge for **or** to go out and relieve nature.
C048	*-xolo*	bark of a tree	species of small veld plant	peace, quiet, calm, goodwill, forgiveness
C048	*-zwe*	nation, tribe, clan, state	rapidly spreading brain disease	
C051, C055, C063, C065, C084	*-bika*	report, announce	species of ant with white spots	
C057, C058, C060, C074, C086	*-zwa*	to perceive with the various senses and thus hear, listen, perceive, understand, taste, smell, sense, feel	sense of feeling, mark of pressure, deep hole, nerve	main rib or spinal of leaf
C070, C075	*-shona*	sink, go down, disappear, go out of sight	set of sun, go down of moon, and even die,	lose heavily, become poor, bankrupt, ruined

The noun *-su* in topics C055 (Table A10.1 and Table A10.2) and C056 (Table A11.1 and Table A11.2) has different meanings when the three different class prefixes are added. When the class prefix *ili-/ama-* is added, *-su* means "plan, method, scheme", or "ring on a cow's horn marking each calf's birth". However, with *isi-/izi-* added, *-su* becomes "stomach, belly, abdomen and appetite" and even 'pregnancy'. When *u(lu)-/izin-* is added to the word, it becomes "human stomach", "tough pliable object" or "soft goatskin cloak worn by women".

In experiments conducted in previous test runs, it was found that the syn-structure of the queries as above described, manages homonyms quite well (Cosijn et. al, 2002a, 2002b, 2002c, 2002d).

From the examples in Table 5.11, it can be seen how frequently mistranslated query keywords occur in s-gram matching.

Table 5.11 *Mistranslated words as found in several queries*

CLEF topic	Source word in text	Matched keywords	English meaning
C041 (Table A1.2)	*ekudleni*	*-nkulukundleni* *unkulukundleni* *onkulukundleni*	species of caterpillar, larva of psychid moth, species of encased grubs
		-nkulukundleni *unkulukundleni* *onkulukundleni*	species of caterpillar, larva of psychid moth, species of encased grubs
		ekuqaleni	in the beginning
C042 (Table A2.2)	*kwezwe*	*-zwezwe* *umzwezwe* *imizwezwe*	spur of a cock
		-zwezwe *umzwezwe* *imizwezwe*	spur of a cock
		-nkweza *inkweza* *izinkweza*	day
C048 (Table A7.2)	*zokuphuma*	*-phuphuma*	overflow, bubble over, be overcome with emotion, have a miscarriage
		-phuphuma *iphuphuma* *amaphuphuma*	species of tree used as a love charm
		-phuphuma *imphuphuma* *izimphuphuma*	overflow, superfluity
C079 (Table A30.2)	*ukuhlolisiswa*	*hlolisa*	cause to live at ease, attack by surprise, ambush

		-wulukuhle *isiwulukuhle* *iziwulukhu*	pouring out in bulk, mass expulsion, abnormally bulky object
		-holisaka *uhholisaka* *ohholisaka*	large sack, mealie bag

Table 5.11 also indicates that in Table A1.2 (Topic C041), the Zulu word *ekudleni* is matched with words that are nowhere similar to the same meaning. This is because *ekudleni* is formed from the word *-dla*, which means food, eating or feast. The word is formed as follows:

uku- + -dla+ -ini

The *uku-* becomes *eku-* because of the influence of *i-* from *ingane* (babies). Also, because of vowel coalescence *dla+ini* becomes *-dleni.*

In Table A2.2 (Topic C042), the word *kwezwe* is formed by *kwa + izwe*. This is not reflected in the dictionary translation process, which means that words in the running text are matched to the available keywords in the Zulu list in base form. The resulting matches differ from the original word *-zwe*, which actually means country/land, where *kwezwe* becomes "the country of".

The same occurs in Table A7.2 (Topic C048), where *zokuphuma* is matched to the shown dictionary entries. The keyword however, is formed through vowel coalescence (Section 5.4.2.3) from *za + uku + phuma*, and thus are not matched to the original word. The stem *-phuma* means "to exit, go out", and thus bears no relevance to the matched dictionary meanings indicated in the above table.

Another example is found in Table A30.2 (Topic C079) with the word *ukuhlolisiswa.* In this instance the match is made to *hlolisa* (attack, surprise), *-wulukuhlu* (abnormally bulky object) or *–hholisaka* (large sack). But, the real stem is derived from *-hlola*, which means "inspect, look into, explore, examine, test".

If the translation for inauguration is analysed (Table A25.1), an interesting and important observation is made:

Table 5.12 *Comparing the dictionary and mother tongue translation of 'inauguration'*

Original translation	*ngokuvulwa*	opening
Mother tongue translation	*ngokugcotshwa*	anointment

In the instance of the mother tongue translation, it seems as if a completely incorrect translation were made, but when placed into context, it is clear that the translator did not take into consideration the context of the sentence, but rather the word itself. For the Zulus, an inauguration is usually associated with royalty. Only kings and queens are inaugurated and anointed as holy ones. For them it might be confusing to anoint a tunnel, but it is the translation for inauguration.

The word *ukugcotshwa* is also an example of word inflection (see Section 5.4.2), where the *-b-* became a *-tsh-*. In this instance, no match could be found, because the system will attempt to match *-gcotshwa*, when the noun stem is actually *-gcoba*.

More examples can be found in Tables A3.1 and A3.2 (CLEF topic C043); Tables A4.1 and A4.2 (CLEF topic C044); Tables A12.1 and A12.2 (CLEF topic C057); Tables A14.1 and A14.2 (CLEF topic C059); Tables A25.1 and A25.2 (CLEF topic C074); and Tables A29.1 and A29.2 (CLEF topic C078).

In Table 5.13, the resulting matches for the Zululised *amakhemikheli* (no such word exists in the dictionary) is shown. However, the Zulu word for chemical that do appear in the dictionary is *-phathalene* or *-thakiweyo*, which are actually adjectives of the word.

Table 5.13 *The Zululised amakhemikheli and the (incorrect) resulting matches in the text*

CLEF topic	**Source word in text**	**Matched keywords**	**English meaning**
C041	*amakhemikheli*	*kheli ikheli amakheli*	address of letter, mannerism, private character
		-kheji ikheji amakheji	cage, bird cage, transport cage
		-kheli umkheli abakheli	person who addresses a letter

To illustrate the process of matching homonyms to their base forms as found in the word index, the following example found in CLEF topic C076 (see Appendix A, Table A27.1 and Table A27.2) can be considered.

A searcher poses the following question:

"In what applications are solar power energy used or considered for future use?"

The typical search terms would be 'applications', 'solar', 'power', 'energy', 'considered', and 'future'. The remainder of the words will be discarded as stop words (see Appendix B for a complete stop word list). As no morphological parser was used in this study, let us assume that an English morphological analyser would normalise this to 'application', 'solar', 'power', 'energy', 'consider', and 'future'. The simple dictionary matches would provide the following information as listed below:

Table 5.14 *Proposed dictionary matches without a morphological analyser*

English	English base form	Zulu *
applications	application	*isithobo*
solar	solar	*-elanga*
power	power	*amandla*
energy	energy	*amandla/isidlakadla*
considered	consider	*cabanga*
future	future	*isikhathi esizayo*

* This is only a simplistic view, and does not take into account that more than one match would be made to the Zulu word.

For the next step in the process, the translated Zulu query has to be matched to the Zulu document. There are various options for matching, depending on the nature of the inverted index. This will be discussed as follows:

5.4.3.1 *Matching the exact forms of the individual words, as they appear in the running text.*

In the first instance, where the inverted index contains the exact forms of the words as they appear in the running text and the matching is simply done on the Zulu singular plus prefixes in the instance of nouns, inflected forms will be missed. For example, in the instance of 'application', the Zulu plural *izithobo* in the document will

not be matched to *izithobo* in the query. The same can be applied to verbs (e.g. -*cabanga*) used in particular contexts as indicated in Table 5.15.

Table 5.15 *Matching inflected verbs in different forms*

Inverted verb	English equivalent
kade ***ngikucabanga***	I have just been thinking about you.
kade ***ngicabanga*** *ngakho*	I have long been thinking it over.
kuphumelela ***abacabangayo*** *emfudweni*	Those who use their brains in study, get on.
wavuka wangasebenza ***ecabanga*** *ukuthi yiSonto*	When he woke he did not work thinking it was Sunday.

Every inflected verb shown in the above table will have problematical results in exact matching. This is because the inflected form is matched to a base form.

5.4.3.2 *Matching the limited normalised word form*

In the second instance, it might be possible to remove a number of prefixes or suffixes through a simple stemming procedure in a simplified morphological analyser. One particular problem that may occur with this particular approach is the difficulty in recognising prefixes, since prefixes may phonetically change because of adjacent letters. Another problem may occur with verbs. This problem can become much worse, due to the complexity of verbal inflections. For example, the middle syllable of the stem -*cabanga* is a bilabial *b* and in the passive mood (through palatalisation), it changes to *tsh* (Doke et al., 1990). The various tenses then become: -*catshangwa* (is considered), -*catschangile* (was considered) and *zokucatshangwa* (will be considered).

Another example is -*sebenza*, where the middle syllable of the stem is also a bilabial *b*. The various tenses in this instance then become: -*setshenziswa* (is used), *setshenziswile* (was used) and *zokusetshenziswa* (will be used).

5.4.3.3 *Matching the normalised inverted index to the dictionary entries*

The third approach is normalising the inverted index to the dictionary entries through n-gram matching between the text and the dictionary entries. This would imply that the search words in the source language are translated and normalised to the stem of nouns and verbs as is found in, for example, the Zulu dictionary of Doke

et al. (1990). Since the inverted index is similarly normalised, there should be a one-to-one match between search terms and inverted index items. The normalised forms of the Zulu translation in the example above then become:

Table 5.16 *The effect of normalisation*

Normalised English	Zulu translation	Normalised Zulu stem
application	*isithobo*	*-thobo*
solar	*-elanga*	*-elanga*
power	*amandla*	*-ndla*
energy	*amandla/isidlakadla*	*-ndla/-dlaka*
consider	*cabanga*	*cabanga*
future	*isikhathi esizayo*	*-zayo*

However, normalising by removing the prefixes and suffixes to reduce nouns and verbs to stems might result in ambiguity, as in the following instance:

Table 5.17 *Normalisation and ambiguity*

Zulu stem	Ambiguous Zulu matches	English meaning
cabanga	*cabanga* *ulu-cabanga*	think, reflect, suppose, consider, imagine cartilage at the end of the breast-bone

In the Zulu dictionary (Doke et al., 1990), singular forms, plural forms and stems for nouns are listed. In using this particular dictionary, ambiguity problems for nouns should be reduced. Verbs, however, are listed only as stems. In addition to possible semantic ambiguity in the stems of the search terms (as in the example of *-cabanga* above), it is also conceivable that n-gram matching between the running text and the dictionary might result in multiple possible matches between a specific word and different dictionary entries. It also yields a very low matching rate in the instance of complex inflected forms. One method for possible resolving this particular problem, would be to take the two highest-ranking items in the inverted index. Another approach would be to use multiple query words and structured queries (Cosijn et al., 2002a, 2002c, 2002d).

5.4.4 Prefixing and suffixing

5.4.4.1 Forming of locatives

The locative indicates place and can, according to the context in which the word appears, be translated as "at...", "in..." or "to...". There are three ways how the locatives of nouns are formed:

- By means of the prefix *ku-*:
  kubaba (to father), *kumfana* (at the boy)
  In these examples, the initial vowel is discarded and substituted by *ku-*.
- By means of *e-...-ini*:
  Umuthi > *emthini* (at/to the tree), *intaba* > *entabeni* (at/to the mountain).
  In these examples the initial vowel is replaced with *e-* and a suffixing *-ini*.
  When the noun ends in *-bu, -bo, -phu, -pho, -bhu, -bho, -mu, -mo, - mbu, -mbo, -mpu* and
  -mpo, the sounds change because of palatalisation as described in Section 5.4.2.1.
- By means of the prefix *e-* without the suffix:
  A number of nouns only take the *e-* in the locative without the suffix *-ini* being added. For example *eMpumalanga* (east), *ekhanda* (the head), *ekhaya* (a home), *esitolo* (a shop), *enkantolo* (magistrate's office), *ehlobo* (summer), *emini* (a day) or *eGoli* (Johannesburg).

Approximately 3% of the errors that occurred in the analysed query words are because of the forming of locatives (see Figure 5.2).

5.4.4.2 Forming of conjunctives

The adverbial formative *na-* may be translated to "with, together with" and is prefixed to the noun. Vowel coalescence takes place as described in Section 5.4.2.3. However, in certain instances *na-* (and, also) is used to join together two or more nouns in Zulu, and as such it coalesces with the succeeding word without altering the grammatical significance of the word. These two forms of *na-* should not be confused with each other.

2% of all the errors that occurred can be ascribed to the forming of conjunctives. In Figure 5.2 this is compared to the other analysed problems as encountered in the text.

5.4.4.3 *Verbal extensions*

Verbal extensions are one of the main characteristics of the verb in Zulu (Bosch and Taljaard, 1988). A verb stem may assume a whole series of different meanings just by suffixing a particular verbal extension.

The different verbal extensions are indicated in Table 5.18.

Table 5.18 *The different verbal extensions occurring in Zulu*

Type of verb extension	Zulu verb	English meaning	Zulu verb with extension	New English meaning
Applied (-el-)	-pheka	cook	-phekela	cook for
Reciprocal (-an-)	-thanda	love	-thandana	love each other
Causative (-is-)	-funda	learn	-fundisa	teach
Neuter (-ek-)	-dla	eat	-dleka	can be eaten
Passive (-w-)	-bona	see	-bonwa	be seen
Passive (-iw-)	-dla	eat	-dliwa	be eaten
Passive	-loba	write	-lotshwa[1]	be written
Passive	-sebenza	work	-setshenzwa[2]	do work
Intensive (-isis-)	-bambisa	hold	-bambisisa	hold very tight
Reduplicated applied (-elel-)[3]	-bona	see / look	-bonelela	look after / care for
Reduplicated verb stem	-hamba	walk	-hambahamba	walk a little
Reduplicated verb stem	-akha	build	-akhayakha[4]	build a little
Reduplicated verb stem	-dla	eat	-dlayidla[5]	eat a little

1 See Section 5.4.2.1 for rules of palatalisation.

2 In certain instances of palatalisation, the extension *-w-* does not immediately follow the consonants.

3 The meaning does not always reflect the correct meaning, but usually indicates an intensified action. Apart from *-elel-*, the extension *-ezel-* is also used to indicate a repeated action.

4 In the instance of disyllabic vowel stems an *-y-* is placed between the stems.

5 In the instance of monosyllabic stems an *-yi-* is placed between the stems.

At least 5% of the problems that were experienced were due to verb extensions (see Figure 5.2 for a breakdown of all types of errors encountered).

5.4.5 Mismatching

5.4.5.1 The enclitic

An enclitic is a formative that is neither a proper suffix, nor used independently. In Zulu there are three formatives considered to be enclitics: *-ke* (then), *-bo* (express insistence) and *nje* (just, merely, simply). These enclitics are suffixed to words (mostly verbs) and have a definite syntactical meaning.

5.4.5.2 The interrogative

Interrogatives may either be in adverbial forms or a suffix. The two forms of interrogative suffixes, as can be seen in the examples in Appendix A, is *-ni* (what?) and *-phi* (where?).

5.5 Conclusions and research findings

In taking this particular study into consideration, it must be emphasised that the translations were done only because no Zulu databases were available. It is indicated in Appendix A that the official translations differ significantly from the mother tongue translations, and it has considerable implications for CLIR and future research on CCIR. This particular study demonstrates how the translations (performed by the mother tongue translators) reflect the local culture of the translators–and how *they* interpret the text. The implications of this for CLIR on Zulu databases are evident. If it is assumed that the content of such databases are IK-related, it is suggested that the 'oral form of Zulu be used to capture the IK content, and not 'academic' Zulu. To overcome the cultural and linguistic barriers experienced in the translation process, the cooperation of native speakers are required to see to it that the original intended message is conveyed in a clear and accurate manner.

Furthermore, through the analysed query words and identified errors (Chapter 5), it was deduced that the Zulu translation aims to transfer every textual aspect. This includes the raw informational aspect, the emotional appeal (quite evident in Zulu) and definite cultural differences in terms of the English translation. This clearly emphasises the difference in communication patterns between Zulu and English.

The three different approaches towards matching the translated Zulu queries to the dictionary entries were outlined in Section 5.4.3.1 – Section 5.4.3.3. These approaches will have to be empirically tested on a corpus to establish which option provides better results. The possibility of applying a simplified morphological

analyser depends largely on the predictability of the use of prefixes and suffixes in the indigenous South African languages. This may differ for the various languages. Promising research are done in developing such a simplified parser for each of these languages, but for purely realistic reasons one may have to opt for non-normalisation in the inverted index of the target language documents (Cosijn et al., 2002a). This may result in low hits for the n-gram matching.

By combining content and metadata searching it might be able to reduce the number of irrelevant items retrieved (Cosijn et al., 2002a). The reason for this is twofold:

- First, by describing the content in detail and capturing additional properties of the content, it could result in a significant drop in the irrelevant items retrieved.
- Second, metadata can act as a filtering mechanism for the identification of the broad field of relevance for instance a search for a certain plant could be delimited to the field of agriculture, medicine or food.

5.6 Chapter synopsis

This chapter focused on the different problems experienced when conducting the empirical studies as described in Chapter 4. One of the main problems associated with dictionary-based CLIR is the abundance of mistranslated query keywords in CLIR queries (also known as *translation ambiguity*). This was evident in the conducted experiments discussed in Section 4.3 by Cosijn et al. (2002a, 2002b, 2002c, 2002d) on Zulu-English queries.

The problems experienced with the dictionary-based approach were discussed in detail. These problems were divided into two main categories–dictionary problems and translations problems, each with its own sub-categories of errors. Each of the mentioned categories and sub-categories of errors have been discussed in detail, as it formed an integral part to this study in terms of the detailed error analysis. Apart from the two categories (dictionary and translation problems), the errors can further be divided into two other categories–those concerned with problematic matching, and those concerned with culture-related issues. These two categories connect with the research questions answered in Chapter 6.

The chapter concluded with a summary on the most important research findings that will lay the foundation for future research in the CCIR field.