Compiling Specialised Glossaries: A Case Study of English-isiNdebele Medical Terms

Nomsebenzi Malele
Department of African Languages, University of South Africa
malelnj@unisa.ac.za

Abstract

The inability to access health care because of a language barrier is a matter of concern which this study seeks to address. There is a critical need to access health care communication by amaNdebele, an ethnic group who speak Southern Ndebele known as isiNdebele, one of the official languages of South Africa. IsiNdebele belongs to the Nguni family, alongside isiZulu, isiXhosa and Siswati. The study examines the compilation of a specialised English-isiNdebele glossary of medical terms. It investigates the methodologies and challenges that are involved in the creation of bilingual medical terminologies for English and isiNdebele. A corpus-based approach together with the prescriptive and descriptive lexicographic methods was employed. Medical corpora were compiled, and consultation with the isiNdebele speaking medical community also took place. This was done so that medical terms which are culturally appropriate could be developed. The findings of this study reveal notable gaps in the medical terminology of isiNdebele. The findings further emphasise the significance of term formation in which the medical community (nurses and doctors) and language (isiNdebele) experts are involved. This study contributes to the broader discourse on equity in healthcare. It also contributes to the intellectualisation of African languages in post-apartheid South Africa.

1 Introduction

Owing to the democratic dispensation of South Africa, the constitution mandated that twelve languages be recognised as official. As much as this has created opportunities, it has created challenges also. The challenges are impacting the development of languages particularly in specialised domains like medicine, which is the focus of this research. This study is on the compilation of an English-isiNdebele glossary of medical terms. It also examines the methodological, theoretical, as well as the practical considerations that are involved in bridging the linguistic gaps in healthcare communication, where isiNdebele dominates.

The compilation of specialised glossaries represents a critical convergence between practical language planning and lexicographic theory especially in cases where indigenous languages like isiNdebele require the development of technical vocabulary.

IsiNdebele belongs to the Nguni language group alongside isiZulu, isiXhosa, and Siswati. Although isiNdebele has official status, just like the mentioned Nguni languages of South Africa to which it belongs, this language still remains marginalised, particularly in specialised domains.

Taljard and de Schryver (2002) maintain that during the last decade, South African languages such as isiNdebele have been reduced to writing whereas languages such as isiZulu have a relatively long literary tradition. IsiZulu has more medical terminology than isiXhosa, followed by isiNdebele and Siswati. That is why the researcher compared the isiNdebele medical terminology with that of isiZulu.

The size of medical terminology creates barriers to effective healthcare delivery for amaNdebele. The development of medical terminology for isiNdebele is not just a matter of lexicographic exercise, but a matter of linguistic justice and healthcare equity as well. The research question this study seeks to answer is: *How can specialised medical glossaries be effectively compiled to serve both the linguistic preservation and practical healthcare communication needs, in the isiNdebele-speaking community?*

The significance of this study is not only confined to the compilation of terminology, but it

encompasses broader issues such as the standardisation of healthcare information, the cultural sensitivity in medical discourse, and the intellectualisation of African languages.

2 Literature Review

According to Cabré's (1999) Communicative Theory of Terminology, terms function within specific discourse communities. This theory provides a foundational understanding of how this is realised. According to this theory, the meaning of terminology is determined by the context. Medical terminology is not exempted from this, therefore, cultural appropriateness and precision must cohabit.

According to Temmerman's (2000) socio-cognitive approach, terminological units are shaped by both cognitive and cultural factors. This approach acknowledges that there may be no direct equivalents for certain medical concepts, therefore innovative ways of terminological creation should be sought.

Although the development of specialised terminology in African languages has received increasing scholarly attention, limited research exists on the development of specialised terminology in isiNdebele. Skhosana's (2009) general lexicographic work on isiNdebele served as foundational principles. Mahlangu's (2014) on the other hand focused on isiNdebele language planning.

This research accentuated the need for the development of domain specific vocabulary in the medical field. This is because the isiNdebele medical terminology remains largely unexplored, which is the gap this study is addressing. Madiba's (2001) research on South African languages' scientific terminology served as significant example for the multilingual terminological development.

Angelelli (2008) conducted research on medical terminology and interpreting, emphasising the importance of precise and error free terminology. This kind of terminology ensures desirable treatment outcomes and patients' safety. Karliner et al. (2007) also conducted research in which they demonstrate that language barriers in healthcare can lead to medical errors which may lead to compromised quality of care.

In their research on language barriers in public healthcare, Benjamin et al. (2016) discovered that patients' inability to communicate effectively with healthcare providers in their mother tongue, significantly impacts health outcomes. This they established in the South African context. The present research emphasises the significance of medical terminology in isiNdebele which will be accessible to amaNdebele.

Kilgarriff and Tugwell's (2001) work confirms significant development in corpus linguistics. Their use of sketch engines which showed collocation patterns, helped in identifying terminological patterns and contexts of usage.

Unfortunately, these methodologies introduce distinct challenges that need the adaptation of methodology when it comes to the implementation of under-resourced languages like isiNdebele.

For the identification and validation of terminology, the corpus-driven approach plays a significant role. Modern lexicographic practice heavily relies on methodologies which are corpus-based. Bowker and Pearson (2002) worked with specialised corpora and established methodological standards for the work of terminology which is corpus-based. Their emphasis was on the compilation of a domain-specific corpus and frequency-based term extraction. This provides methodological guidance for this present research.

3 Theoretical Framework

This study operates within a multidisciplinary theoretical framework that integrates insights from terminology science, lexicographic theory, and language planning. The framework is anchored by two primary theoretical constructs:

3.1 Communicative Theory of Terminology

Cabré's (1999) Communicative Theory of Terminology serves as the primary theoretical foundation, emphasising that specialised terms derive their meaning from communicative contexts rather than existing as isolated semantic units. This perspective is particularly relevant to medical terminology compilation, where terms must function effectively within specific healthcare communication scenarios.

3.2 Language Planning Theory

Cooper's (1989) comprehensive treatment of language planning provides the sociolinguistic framework for understanding terminology compilation as a form of corpus planning. This perspective situates the development of isiNdebele medical terminology within broader language planning objectives, including language standardisation, intellectualisation, and status enhancement.

4 Research Methodology

4.1 Research Design

This study employs a mixed-methods approach combining corpus-based quantitative analysis with qualitative consultation processes. The research design integrates descriptive lexicographic methods for identifying existing terminological patterns with prescriptive approaches for developing new terminological solutions where gaps exist.

The methodological approach is inherently participatory, involving isiNdebele-speaking healthcare professionals, isiNdebele interpreters and isiNdebele subject advisors in the terminology development process. This community-driven approach ensures cultural appropriateness and practical applicability of the compiled glossaries.

4.2 Data Collection

4.2.1 Corpus Development

A specialised medical corpus was compiled from multiple sources to provide the empirical foundation for terminology identification. The corpus is made up of the following:

- IsiNdebele medical information for patients.
- Transcribed conversations between isiNdebele-speaking patients and isiNdebele-speaking healthcare providers.
- Training materials for isiNdebele-speaking health workers.
- Government health policy documents translated into isiNdebele.

Regarding copyrights for the medical pamphlets, the process was daunting. It was not clear who holds copyright, this is confirmed by Bowker and Pearson (2002) when they say that the process of getting copyright permission is not a straightforward one. Regarding the transcription of this research spoken corpus, it was done by the researcher herself.

4.2.2 Expert Consultation

Semi-structured interviews were conducted with key stakeholders including medical professionals fluent in both English and isiNdebele (n=12), Parliament of South Africa isiNdebele interpreters (n=2) and isiNdebele-speaking nurses (n=10), and two isiNdebele-speaking medical doctors.

These consultations provided expert validation of terminological choices and cultural insights essential for appropriate term development.

4.3 Data Analysis

4.3.1 Corpus Analysis

For term extraction as well as frequency analysis, computational tools were employed. The AntConc concordance program, 4.3 was used. AntConc is a freeware corpus analysis toolkit which is used for concordancing and text analysis. This program is provided under a licence which is on Windows. It is from translation and language software. It has absolutely no restrictions on usage. As mentioned, this PC software is free, both its download and installation. The WordList, KeyWordList and Key Word In Context of the AntConc tool was used

WordList

This tool counted all the words in the medical corpus and presented them in an ordered list. It enabled the researcher to quickly recognise the most frequent words in a corpus. The WordList shows the terms' rank, frequency (that is, how many times the word appears) and the range of the term (the number of files the particular word appears in).

KeyWord List

Through this tool words which are unusually frequent (or infrequent) were identified in the analysis corpus (which is technical and domain-specific) in comparison with the same words in the reference corpus (which is non-technical and non-domain specific). The running corpus (RC) had 189 517 running words. The size of the RC ensures that the content of the subjects covered is diverse. The analysis corpus had 96 732 running words. This corpus comprises of a variety of medical topics. The size of the corpus provides sufficient data for frequency-based term extraction while remaining manageable for detailed qualitative analysis.

The calculations of Keynes were done in order to recognise the most significant medical terms that needed to be included in the glossary. Those terms that were occurring with frequencies above thresholds that are predetermined, were the ones selected for thorough analysis.

Key Words In Context/ Concordance Tool

This helped with the creation of an in-depth knowledge of the usage of terms in an acceptable way. It enabled the researcher of the current study to find patterns of similarity or contrast in the words surrounding the search term, for example, a term such as *udorhodere* 'doctor' was inserted to be searched.

Through this program, high-frequency medical terms were both identified and analysed in terms of their contextual usage patterns. Collocation analysis helped to reveal terminological patterns and to identify gaps that are in the existing terminology.

4.3.2 Terminological Analysis

Each identified term underwent detailed terminological analysis: In the following paragraphs the isiNdebele equivalents are analysed for instance in terms of what term-formation strategy has been used. See the following:

- **Existing isiNdebele equivalents or related terms**, for example for the English term 'pap-smear', an isiNdebele equivalent existed which is *ukuhlolwa komlomo wesibeletho*. This is translating by paraphrasing, and it literally means 'the examining of the mouth of the cervix'. Paraphrasing is translating by giving a short description or explanation (Taljard, 2008). The English term 'epilepsy' for instance has an existing isiNdebele equivalent which is *isithunthwana* which was created from seemingly nothing.
- **Cultural appropriateness and sensitivity considerations.** According to this sub-topic, culture shapes how terms are formed. It is a fact that there are words that according to culture are deemed sensitive and inappropriate. Instead of such words a more polite one is used. The words that are deemed inappropriate are considered taboo and impolite. Such words are embarrassing to talk about. These are words about sex, private body parts, death, to mention but a few. For example, the isiNdebele equivalent for the English word 'miscarriage' is *ukubuya endleleni* which literally translates as 'to come back from the road'. It is a euphemistic word, used instead of *ukuphunyelwa sidisi* which is deemed insensitive and vulgar according to the culture. Both these isiNdebele examples are paraphrased. The term 'to die' is translated as *ukulala* 'to sleep' or *ukukhamba* 'to go' to avoid *ukufa* which is deemed insensitive.

  The above examples also talk to the register considerations as well as usage contexts. Register refers to how language is used based on social context and audience. Choice of words is evident in the isiNdebele spoken corpus more than its written counterpart, particularly in the case of terms that are regarded as vulgar and inappropriate (as mentioned in the aforegoing paragraph). For instance, sexual intercourse is referred to as *ukuya emsemeni* in the written corpus. In the spoken corpus it is referred to as *ukulalana* 'to sleep with each other'*, ukukha umrorho* and *ikonzo yabantu abadala*; the very act is called *ukuya emsemeni* which loosely translates as 'to go to the reed-mat' or *ukukha umrorho* which loosely translates as 'to pick up the vegetables' or *ikonzo yabantu abadala* which loosely translates as 'old people's service or fellowship'. This is euphemistic. Spoken language is characterised by more euphemism as compared to its written counterpart (Malele, 2021).

- **Morphological and phonological adaptation requirements.** Mahlangu (2007) argues that in the case of isiNdebele, the loan words from English that have consonant clusters such as /sl/, /sq/, /sp/, /st/ and /sch/ insert a vowel. See the following example: The English term 'scanner' is written as *isikena* in isiNdebele (Mahlangu, 2007).

4.3.3 Validation Procedures

Multiple validation procedures were implemented to ensure the quality and appropriateness of compiled terminology. Since human beings will always be final judges in any terminological work, expert reviews by medical and linguistic professionals were implemented. IsiNdebele speaking doctors and isiNdebele Curriculum Implementers (for Further Education and Training) assisted with the verification of terms. For instance, most isiNdebele translations refer to the English 'groin' as *imbilapho*. The experts maintain that *imbilapho* is actually a 'lymph node' in English and not a 'groin'. A 'groin' is not a medical condition, but a body part.

5 Findings and Discussion

5.1 Terminological Gaps and Challenges

The corpus analysis revealed significant gaps in isiNdebele medical terminology, particularly in specialised medical terms. Approximately 70% of identified high-frequency medical terms lacked established isiNdebele equivalents, necessitating new term creation or adaptation. The gaps in the isiNdebele medical terminology were identified by comparing the isiNdebele terms with the isiZulu basic medical terms available at http://hdl.handle.net/10386/3516. The researcher saw it fit to compare isiNdebele

with isiZulu, as it is one of the most developed languages in the Nguni family, to which isiNdebele belongs.

These gaps reflect the historical marginalisation of isiNdebele in formal medical education and practice, creating barriers to effective healthcare communication.

5.2 Terminological Development Strategy

To address the identified terminological gaps, morphological adaptation was used, as a strategy:

The morphology of isiNdebele is agglutinative. Transliteration has proven to be the most preferred method of term formation, amongst amaNdebele. The transliterated term *ihayibhladi* for 'high blood pressure' is used more than its coined counterpart, which is *isigandelelo seengazi*. Following transliteration is paraphrasing. It is the second most preferred method of term formation used by amaNdebele when creating terms. For example, the English term 'dentist' is referred to as *udorhodere wamazinyo* (doctor of teeth) and 'dermatologist' as *udorhodere wesikhumba* (doctor of skin).

The third mostly used term-formation method in isiNdebele is compounding. The challenge with compounding has been the issue of hyphenation. There is a lot of inconsistency as far as compounds are concerned. Mahlangu (2013) maintains that the problem with isiNdebele compound nouns is the issue of hyphenation. Mahlangu (2013) says that some compounds are hyphenated, and some are unhyphenated for no apparent reason. The isiNdebele term *umbulalasihlangu* (AIDS) for instance, is sometimes written as *umbulala-sihlangu* (with a hyphen) and in other instances as *umbulalasihlangu* (without a hyphen) (Mahlangu, 2013).

6 The Study's Contributions

The combination of corpus-based analysis with health community consultation provides a balanced approach that ensures both empirical grounding and cultural appropriateness and acceptance. The validation procedures developed for this study provide templates for quality assurance in specialised terminology compilation for African languages.

The compiled glossaries provide immediate practical resources for healthcare communication in isiNdebele-speaking communities. These materials can support medical interpreter training programs, healthcare provider competency training, and community health worker capacity building.

Ikhwezi National Language Board was established to carry the mandate of isiNdebele terminology development. Although there are terminology lists that have been approved by the board, there is no list on isiNdebele medical terms. This work will contribute and speed-up the development of isiNdebele medical terms and also assist this board in employing computational methods.

The research findings have significant implications for language policy in healthcare settings: They underscore the importance of including indigenous languages in medical education curricula and requirements for healthcare interpreting services in indigenous languages.

7 Future Research Directions

This study opens several avenues for future research: Expansion to other South African indigenous languages, the development of multimedia terminology resources and investigation of terminology adoption and standardisation processes.

8 Conclusion

The compilation of a specialised English-isiNdebele glossary of medical terms promotes equitable healthcare access while it addresses historical marginalisation. The research findings reveal both the possibilities and the challenges that are fundamental in developing specialised terminology for indigenous South African languages particularly isiNdebele, as it is the focus of this paper. While there are significant gaps that exist in isiNdebele medical vocabulary, currently such gaps are gradually bridged by the fact that the isiNdebele speaking medical professionals have the innovative capacity to provide powerful foundations for terminological development.

The theoretical framework developed for this study, combining communicative terminology theory with the socio-cognitive approach, acknowledges that even if there are no direct isiNdebele equivalents for certain medical concepts, innovative ways of terminological creation should be sought and implemented. Language planning approaches offer a vigorous foundation for similar research in other African language contexts.

This research is a testament to the fact that when isiNdebele and other indigenous languages

are given the necessary developmental support, such languages can actually serve a specialised discourse function. This research also contributes to the broader project of African language intellectualisation. The compiled glossary represents concrete resources for improving healthcare communication.

The development of specialised terminology in isiNdebele represents both a constitutional imperative and a practical necessity for achieving a true equitable healthcare delivery.

The success of this terminology compilation project depends ultimately on institutional adoption and community usage. Future efforts must focus on implementation strategies that ensure the developed terminology becomes integrated into actual healthcare practice, serving the communities whose linguistic heritage it seeks to preserve and develop.

Limitations of the Study

Several limitations constrain the generalisability of this research: Geographic concentration in specific isiNdebele-speaking regions, limited representation of specialised medical sub-domains, time constraints preventing longitudinal validation of terminology adoption and resource limitations affecting corpus size and diversity.

Ethical Considerations

This research received ethical approval from the University Ethics Committee and followed established protocols for community-based research. Informed consent was obtained from all participants, and community ownership of developed terminology was established through formal agreements with representative community organisations.

References

Claudia V. Angelelli. 2008. *Medical interpreting and cross-cultural communication*, Cambridge University Press, Cambridge.

Ereshia Benjamin, Leslie Swartz, Linda Hering and Bonginkosi Chiliza. 2016. Language barriers in health: lessons from experiences of trained interpreters working in public sector hospitals in the Western Cape, South African Health Review, pages 73-81. https://hdl.handle.net/10520/EJC189317

Lynne Bowker and Jennifer Pearson. 2002. *Working with specialized language: A practical guide to using corpora*, Routledge. DOI:10.4324/9780203469255

Maria Teresa Cabré. 1999. *Terminology: Theory, methods, and applications*, John Benjamins, Amsterdam/Philadelphia. DOI: https://doi.org/10.7202/004006ar

Robert L. Cooper. 1989. *Language planning and social change*, Cambridge, Cambridge University Press. https://doi.org/10.1017/CBO9780511620812

Leah S Karliner, Elizabeth A Jacobs, Alice Hm Chen and Sunita Mutha. 2007. Do professional interpreters improve clinical care for patients with limited English proficiency? A systematic review of the literature, *Health Services Research*, 42(2):727–754. doi.org/10.1111/j.1475-6773.2006.00629.x

Adam Kilgarriff and David Tugwell. 2001. WORD SKETCH: Extraction and display of significant collocations for lexicography. In *Proceedings of the ACL Workshop on COLLOCATION: Computational Extraction, Analysis and Exploitation*, pages 32-38. Toulouse, Association for Computational Linguistics.

Mbulungeni Madiba. 2001. *Multilingual terminology development in South Africa*. Johannesburg: University of the Witwatersrand Press.

Katjie S Mahlangu. 2007. Adoption of loan words in isiNdebele. Unpublished M.A. Mini Dissertation, Pretoria, University of Pretoria.

Katjie S Mahlangu. 2013. Challenges to the usage of a hyphen in compound nouns in Southern Ndebele, *G.S.T.F. International Journal on Education*, 1(1):107–111.

Katjie S Mahlangu. 2014. Language planning and policy in post-apartheid South Africa: The case of isiNdebele, *Language Matters*, 45(2):223–240.

Nomsebenzi J Malele. 2021. The use of corpora in the compilation of a specialised English-isiNdebele glossary of medical terms. Unpublished DLitt et Phil thesis, Pretoria, University of South Africa. https://hdl.handle.net/10500/28164.

Philemon B. Skhosana. 2009. *The linguistic relationship between Southern and Northern Ndebele*, Pretoria, University of Pretoria Press.

Elsabé Taljard. 2008.Issues in scientific terminology in African Languages. Paper presented at the CEntrePol workshop held at the University of Pretoria on 29 March 2007, Supported by the French Institute of Southern Africa, pp.88-91.https://hal.science/file/index/docid/799553/filename/Cahiers-

Elsabé Taljard and Gilles-Maurice de Schryver. 2002. Semi-automatic term extraction for the African

languages, with special reference to Northern Sotho, *Lexikos*, 12:44-74. https://doi.org/10.5788/12-0-760

Rita Temmerman. 2000. *Towards new ways of terminology description: The sociocognitive approach*, Amsterdam/Philadelphia, John Benjamins. https://doi.org/10.1075/tlrp.3