

Using MonoConc Pro to teach and learn lexical collocations in Xitsonga

Mlambo Respect & Matfunjwa Muzi

South African Centre for Digital Language Resource (SADiLaR), North-West University, Potchefstroom, South Africa

Respect.Mlambo@nwu.ac.za &

Muzi.Matfunjwa@nwu.ac.za

Abstract

Few language resources have been developed for indigenous languages in South Africa. Surprisingly, these are also official languages which constitutionally share the same language status as other languages. One of the major challenges for the development of basic language resources is the lack of digital corpora that can be used to train and develop the resources. Such a challenge has impeded the use of technology for research, learning, and teaching domains in indigenous languages. In this study, we used MonoConc Pro, a concordancer, to demonstrate how language users can utilise the software to display lexical collocation from a corpus for teaching and learning purposes. We illustrated how corpus can be used simultaneously with language technology to teach and learn aspects of linguistics in a form of lexical collocations in Xitsonga. An Autshumato Xitsonga Monolingual Corpus (AXMC) that was retrieved from the South African Centre for Digital Language Resources repository was used as data for analysis. The AXMC is a corpus that was collected and semi-automatically aligned at the sentence level during the Autshumato project. To search for lexical collocations, we interrogated the AXMC corpus using the MonoConc Pro program. A semi-automatic search for collocates of Xitsonga adjectives *lavakulu*, *letikulu*, *lavantsongo*, and *letinene* was conducted. The study found that lexical collocations or words that co-occur with adjectives are nouns, adjectives, possessives, and relatives. It was also observed that each adjective frequently collocates with certain nouns belonging to a specific class. The results obtained suggest a practical way in which language technologies can

be used to explore corpora and examine language patterns for teaching and learning. We hope that this line of study will lead to the use of modern language resources to examine linguistic traits in indigenous languages.

Keywords: MonoConc Pro, Lexical collocations, Xitsonga, Concordancer, Language resources

1 Introduction

There is a significant gap in the development of South African official indigenous languages, particularly in the use of language resources for research, teaching, and learning. Some languages, such as English and Afrikaans, have received preferential attention in terms of development and use of contemporary language resources, whereas indigenous languages with fewer speakers, such as Siswati, isiNdebele, Tshivenda, and Xitsonga, are still in the process of development (Finlayson and Madiba 2002 as cited by Mlambo et al., 2021, p. 82). The absence of contemporary language resources and appropriate corpora required for building digital tools for indigenous languages has hampered research, learning and teaching in these languages using modern technologies. Furthermore, political, cultural, and socio-economic issues have hindered the development and integration of contemporary language resources into indigenous languages (Unesco, 2019). Meanwhile, languages with such resources and corpora have benefited in the research, learning, and teaching domains (Yunus, 2017, p. 297).

There have been some developments in South Africa with the establishment of language research infrastructure, such as the South African Centre for Digital Language Resources (SADiLaR), which provides corpora and builds digital tools for all 11 official languages. This centre facilitates the creation, administration, and dissemination of contemporary language resources, as well as relevant digital tools, which are made publicly available to communities for teaching, learning, and research purposes (<https://www.sadilar.org>). However, other languages are still underserved in some of SADiLaR's contemporary language resources.

Therefore, additional efforts are needed to develop and promote available language resources for all indigenous languages.

The study focuses on lexical word combinations known as collocations in linguistics. Collocations, or the co-occurrence of words, are crucial in language teaching and learning (Jafarpour, 2013, p. 57). We employ Xitsonga corpus, which is a language with a disjunctive writing system, to demonstrate how a concordancer can be utilised to teach as well as to learn Xitsonga lexical collocations. We pay attention particularly to the parts of speech that collocate with adjectives. Barlow's MonoConc Pro was used to illustrate how lexical collocations can be viewed and learned from a corpus using a concordancer. This technique was coined by (John, 1991) as Data-Driven Learning, in which learners or students of a language use electronic corpora to actively investigate a language and its grammatical patterns such as collocations and prepositional colligations (Lewis, 2000).

Concordancers such as WordSmith Tools, Multiconcord, and ParaConc have primarily been investigated in South African languages by scholars such as Taljard and De Schryver (2002), Madiba (2004), Moropa (2007), Ndhlovu (2016), Shoba (2018), and Mlambo (et al., 2021) for identification and extraction of terminology lists for dictionary compilations. Despite the benefits that have been observed, these concordancers have not been employed in teaching and learning grammatical aspects of Xitsonga. Therefore, this study attempts to close the gap in the use of corpus-based methods and concordancers to teach and learn lexical collocation in Xitsonga.

2 Related work

According to Moehkardi (2002), all languages contain words that frequently co-occur with other words(s) in units and their co-occurrences correspond to particular grammatical norms of a specific language, and such words are known as collocations. Benson et al. (1986) classified collocations into two types: grammatical collocations and lexical collocations. Grammatical collocations are made up of a noun, an adjective

or a verb, plus a particle (a preposition, an adverb, or a grammatical structure such as an infinitive, gerund, or clause) (Bahns, 1993). The grammatical collocations are occasionally idiomatic and lexicalised as single units since their meanings do not correspond to the real meanings of the parts. Meanwhile, lexical collocations do not include prepositions, infinitives, or relative clauses, but instead consist of different combinations of nouns, adjectives, verbs, and adverbs (Bahns, 1993). However, in this study we are only interested in lexical collocations.

Pirmoradian and Tabatabaei (2013) researched Iranian English first language (EFL) students. Their study's goal was to examine the impact of using the Collins Collocation Dictionary (CCD) as a concordancer tool on Iranian EFL university students' acquisition of English lexical collocations. A total of 30 students was randomly assigned to one of the two groups: the experimental group and the control group. The experimental group was tasked with identifying miscollocations in 10 lexical collocations. Meanwhile, the control group was instructed to identify lexical collocations in the texts they were given. Pirmoradian and Tabatabaei (2013) discovered that using the CCD as a concordancer had a substantial impact on the subjects' overall performance of students in learning English lexical collocation. As CCD was used as a concordance tool, students' results on lexical collocation improved considerably when compared to those without the concordance tool.

The findings also demonstrated that there is a substantial difference between the impacts of concordance approaches and traditional methods on students when it comes to learning English lexical collocation for Iranian EFL learners. This means that students who utilise a concordancer to learn lexical collocations are more likely to outperform students who learn collocations through traditional methods.

The current study also draws from the research by Jafarpour et al. (2013), who compared the effects of the corpus-based approach with the effects of the traditional approach to learning collocations.

The study was conducted on two groups of English second language (L2) students. The experimental group implemented a concordancer, while the control group used a traditional technique. In terms of proficiency and collocation competency, the individuals in each group were at the same level when the experimentation began. Jafarpour et al. (2013) observed that using a corpus-based approach to teaching collocations and their use in writing is more beneficial for L2 learners than traditional techniques. In general, the findings revealed that using concordancers to teach collocations is more fruitful and efficient than using traditional collocation teaching methods.

From the literature consulted, no study has investigated lexical collocation in Xitsonga either with the aid of language resources or using a corpus-based approach. This reveals the critical need for employing modern language tools to explore grammatical trends in indigenous languages by using corpora. As a result, we believe that our study will pave the way for the use of concordancers and corpora to aid in the teaching and learning of Xitsonga lexical collocations and other linguistics traits respectively.

3 Methodology

This is qualitative research that employs an Autshumato Xitsonga Monolingual Corpus (AXMC) acquired from an online SADiLaR repository as data for analysis. The AXMC is a corpus that was collected and semi-automatically aligned at the sentence level during the Autshumato project. This project was funded by the Department of Arts and Culture and aimed to create tools, resources, and corpora for all South African languages (Groenewald & Fourie, 2009, p. 191).

MonoConc Pro which is a software invented by Michael Barlow was used to analyse the data. The software is utilised as a text-searching program with a user-friendly interface. Like other concordancers, MonoConc Pro has numerous capabilities such as constructing word lists (in both alphabetical and frequency order), creating concordance output, handling huge corpora quickly, providing collocation information, and

operating with tagged or untagged texts (Reppen, 2001, p. 32). The primary purpose of using MonoConc Pro in our study is to demonstrate its utility on how lexical collocations can be viewed from a corpus using the concordancer for purposes of lexical competence in Xitsonga. To display collocations, the AXMC was imported in its text format into MonoConc Pro. Then, the 'search engine' function on the software was utilised to explore collocations of a particular adjective. When the search process was complete, all collocates of the entered adjectives were shown as results.

4 Data presentation and analysis

This section demonstrates how MonoConc Pro can be used to view lexical collocations for teaching and learning outcomes. MonoConc Pro was chosen for its functionality and features that enable us to find and visualise the lexical collocations in a language. The search engine function of MonoConc Pro was utilised to illustrate ways in which lexical collocations could be examined from corpora. A description of the operation of MonoConc Pro and its search engine feature in identifying collocations follows.

4.1 Search engine

The search engine is the main MonoConc Pro feature that was used to interrogate the corpus to view lexical collocations. To search for a specific adjective, we used the search function found on the MonoConc Pro interface bar. Searching for the adjective in Xitsonga resulted in the retrieval of all occurrences of the searched adjective together with its collocates. The target adjective was displayed in blue with the correct collocates highlighted in red on its left side. Using the search engine is beneficial in sorting words and viewing the collocation immediately to the left or right of the target word as this gives us vital information on how the word has been used in the context (Reppen, 2001, pp. 33-34).

Results of the searched Xitsonga adjectives *lavakulu*, *letikulu*, *lavantsongo*, and *letinene* are given in a concordance format in figures. All instances in which the searched adjectives have been used are

presented, and these occurrences are utilised to see all words in the corpus that collocate with the searched adjectives. An adjective is a lexical category that modifies a noun by providing additional information about the referent (Radford et al., 2009, p. 130). Adjectives in Xitsonga are constructed by adjectival concords and adjectival stems. The adjectives that are shown

in this study have adjectival concords *lava-* and *leti-*, which are derived from demonstrative pronouns and adjectival stems *-kulu* (big), *-ntsongo* (small), and *-nene* (good). As illustrations, the results of the lexical collocations for the searched Xitsonga adjectives *lavakulu*, *letikulu*, *lavantsongo*, and *letinene* are presented consecutively in Figures 1, 2, 3, and 4.



Figure 1: Results of the search adjective 'lavakulu'

The adjective *lavakulu* has collocated with the nouns *vanhu* (people) and *vaofisiri* (officers) in Figure 1. These nouns belong to class 2 which is known as the human class that uses the plural prefix *va-*. The queried adjective also collocates

with subject nouns *vuleteri* (guidance) and *dyondzo* (lesson) and their concords *ya* and *bya*. The software was able to display all collocates of the searched adjective. However, some were not highlighted.



Figure 2: Results of the search adjective 'letikulu'

In Figure 2, the searched adjective *letikulu* also collocates with different nouns such as *tikhaphani* (companies), *tindhawu* (places), *tiphurhojeke* (projects), *timhaka* (news), and *tingana* (disgrace) that are categorised in noun class 10. This class uses the plural prefix *ti-*. The figure also shows that

the adjective *letikulu* can collocate with the adjective *timbirhi* (two) that gives numerical information about the nouns *timhaka* (news). It is interesting to observe that the software accurately identified *timhaka* (news) and *timbirhi* (two) as collocates of *letikulu*.



Figure 3: Results of the search adjective 'lavantsongo'

In Figure 3, the words that collocate with the searched adjective *lavantsongo* are similar to those in Figure 1. The noun *vanhu* (people) is the main word that has collocated with the adjective *lavantsongo*. Besides collocating with the noun *vanhu* (people), the adjective *lavantsongo* also collocates

with possessive qualificative *va hina* (of ours). This possessive qualificative was formed by combining the possessive concord *va* (of) and possessive stem *hina* (ours) which is derived from the absolute pronoun. From this figure, we observed that almost all the lexical collocations that involve the searched adjective were highlighted.



Figure 4: Results of the search adjective 'letinene'

From Figure 4, the occurrences show that all words which collocate with the adjective *letinene* are the same as in Figure 2. The collocates *tindlu* (houses) and *timhaka* (news) are nouns that are classified under class 10 in the Xitsonga noun

class system. The remaining collocate *to blaya* (many) that is highlighted in the figure is a relative qualificative. It was formed by a relative concord *to* and relative stem *blaya* (count) which is derived from a verb.

5 Conclusion

In the study, we used MonoConc Pro to explore and visualise collocates of adjectives from the Xitsonga monolingual corpus. We deduce that lexical collocations or words that co-occur with adjectives, namely *lavakulu*, *letikulu*, *lavantsongo* and *letinene*, are nouns, adjectives, possessives, and relatives. It was discovered that each adjective often collocates with particular nouns that belong to a specific noun class system. The adjectives that were studied collocate with nouns in classes 2 and 10 of the Xitsonga nouns class system. The utilisation of the MonoConc Pro software has helped us to realise that adjectives do not only co-occur with nouns but also with other parts of speech. The overall results have demonstrated the usefulness of concordancers such as MonoConc Pro to analyse Xitsonga language patterns such as lexical collocations. The study also exhibits the significance of concordancers and corpora in increasing language proficiency and collocational competence awareness among speakers of the language in question. The use of such language technology provides a practical approach to teaching and learning lexical collocations with the aid of corpora, other than relying on traditional methods of teaching and learning linguistic traits of indigenous languages.

References

- Bahns, J 1993, 'Lexical collocations: a contrastive view', *ELT Journal*, vol. 47, no 1, pp. 56-63.
- Benson, M, Benson, E, & Ilson, R 1997, *The BBI dictionary of English word combinations*, John Benjamins Publishing Company, Amsterdam.
- Finlayson, R & Madiba M 2002, 'The intellectualisation of the indigenous languages of South Africa: Challenges and prospects', *Current Issues in Language Planning*, vol. 3, no 1, pp. 40-61.
- Groenewald, HJ & Fourie, W 2009, 'Introducing the Autshumato Integrated Translation Environment' *13th Annual Conference, EAMT Proceedings 2009*, Barcelona, Spain, May 14-15, 2009, pp. 190-196.
- Jafarpour, AA, Hashemian, M & Alipour, S 2013, 'A corpus-based Approach toward teaching collocation of synonyms', *Theory and Practice in Language Studies*, vol. 3, no 1, pp. 51-60.
- Johns, T 1991, 'Should you be persuaded: Two examples of data-driven learning', in T. Johns & P. King (Eds.), *Classroom Concordancing. English Language Research Journal*, vol. 4, pp. 1-16.
- Lewis, M 2000, 'Language in the lexical approach', in M Lewis, *Teaching collocation: Further developments in the lexical approach*, Language Teaching Publications, Hove, pp. 126-54.
- Madiba, M 2004, 'Parallel corpora as tools for developing the indigenous languages of South Africa with special reference to Venda', *Language Matters*, vol. 35, no 1, pp. 133-147.
- Mlambo, R, Skosana, N & Matfunjwa, M 2021, 'The extraction of terminology list using ParaConc for creating a Quadrilingual dictionary', *Southern African Linguistics and Applied Language Studies*, vol. 39, no 1, pp. 82-91.
- Moehkardi, R 2002, 'Grammatical and lexical English collocations: Some possible problems to Indonesian learners of English', *Humaniora*, vol. 14, no 1, pp. 53-62.
- Moropa, K 2007, 'Analysing the English-Xhosa parallel corpus of technical texts with ParaConc: A case study of term formation processes', *Southern African Linguistics and Applied Language Studies*, vol. 25, no 2, pp. 183-205.

- Ndhlovu, K 2016, 'Using ParaConc to extract bilingual terminology from parallel corpora: A case of English and Ndebele', *Journal of Literary Criticism, Comparative Linguistics and Literary Studies*, vol. 37, no 2, pp. 1-12.
- Pirmoradian, M & Tabatabaei, O 2012, 'The enhancement of lexical collocation learning through concordancing: A case of Iranian EFL learners', *The Modern Journal of Applied Linguistics*, vol. 4, no 4, pp. 185-200.
- Radford, A, Martin, A, David, Herald, C, Andrew, S 2009. *Linguistics: An Introduction*, Cambridge University Press, Cambridge.
- Reppen, R 2001, 'Review of MonoConc Pro and Wordsmith tools', *Language Learning and Technology*, vol. 5, no 3, pp. 32-36.
- Shoba, FZ 2018, 'Exploring the use of parallel corpora in the compilation of specialized bilingual dictionaries of technical terms: A case study of English and isiXhosa', PhD thesis, University of South Africa.
- Statistics South Africa 2012, *Census 2011: Census in brief*, Statistics South Africa, Pretoria.
- Taljad, E & De Schryver GM 2002, 'Semi-automatic term extraction for the African languages, with special reference to Northern Sotho', *Lexikos*, vol. 12, pp. 44-74