

Making sense of *kuningi* using a corpus linguistic analysis

Langa KHUMALO

SADiLaR, North-West University, Potchefstroom, South Africa.

langa.khumalo@nwu.ac.za

Abstract

In 2023 the Pan South African Languages Board (henceforth PanSALB) pronounced the word *kuningi* as the Inaugural South African word of the year for social media. In its statement of 16 October 2023, PanSALB stated in this pronouncement that the word is reflective of “[...] real language use” (Huluhulu, 2023). *Kuningi* is an IsiZulu word literally meaning “it’s a lot”. This study analyses the meaning of *kuningi* using a corpus linguistic approach. Two analysis corpora (AC) were created from well-known social media platforms, namely X (formerly Twitter) and Instagram. The other two analysis corpora (AC) were created from two prominent IsiZulu newspapers, respectively *Isolezwe* and *Ilanga*. The corpus data that was collected for this study was limited to the data produced between 1 January 2023 and 30 November 2023.

The study uses the corpus linguistic approach as a basis for the theoretical analysis. According to Sinclair (2005), a corpus is "a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research." The advantage of such a theoretical approach is that "[...] a corpus [is] stored in a computer, it is easy to find, sort and count items, either as a basis for linguistic description or for addressing language-related issues and problems" (Kennedy, 1998: 11). A corpus is thus a collection of naturally occurring texts derived from real life language use in either written or spoken form, which is then processed, stored, and accessed by means of computers (Khumalo, 2015: 497). Such a corpus is then useful as a basis for, *inter alia*, investigating language use.

The analysis corpora from X and Instagram (henceforth AC_X and AC_I) form the basis for the analysis of the word *kuningi* from a social media perspective, and therefore is viewed as projecting the most understood public meaning of the word *kuningi*. Because of the nature of the text types that are derived from these platforms, the corpus data is small. AC_X has 870 tokens, while AC_I has 4 218 tokens. *Kuningi* is the most frequent word in AC_X as shown in Figure 1. It is the third most frequent word in AC_I.

| Cirrus | | | |
|--------|----------|-------|-------|
| Terms | | Links | |
| | Term | Count | Trend |
| 1 | kuningi | 16 | |
| 2 | 2023 | 15 | |
| 3 | nov | 15 | |
| 4 | replying | 8 | |
| 5 | nje | 5 | |
| 6 | black | 3 | |
| 7 | buying | 3 | |
| 8 | like | 3 | |

Figure 1. Frequency List for the AC_X.

The study uses the two corpora derived for the most popular IsiZulu newspapers *Isolezwe* (henceforth AC_{ISO}) and *Ilanga* (henceforth AC_{ILA}), to comparatively analyse the standard use, and therefore, standard meaning of the word *kuningi*. The two corpora are also limited to the period 1 January 2023 - 30 November 2023. AC_{ISO} and AC_{ILA} respectively have 4 061 tokens and 2 915 tokens. To enable the comparison, the study uses Voyant Tools to query and analyse the two sets of corpora. Voyant Tools is used a lot in digital humanities to facilitate the analysis of textual data using computational tools. Voyant Tools is an open-source, web-based application for performing data analysis. It enables close text reading and interpretation using sophisticated functionalities for analysing the key-word-in-context. Its use enabled the swift creation of the wordlist, concordance analysis and other (syntactic) correlations. The study will show that in the social media context the word *kuningi* has extended its semantic range to include “[...] is hard”, “[...] is difficult” and “[...] is complicated”. While its use in the formal public media has retained its general meaning “[...] many” and “[...] a lot”.

Section 1 briefly discusses the PanSALB and its role in the promotion of the 12 official languages in South Africa. The context of the South African Word of the Year is also discussed in the section. Section 2 discusses crowdsourcing and the corpus. The section discusses in detail the four corpora and how these were carefully collected. We analyse the keyword in Section 3 using a corpus linguistics as a framework, we discuss the findings in Section 4 and conclude in Section 5.

Keywords: Comparative Analysis, Corpus, Corpus Linguistics, *Kuningi*, Voyant Tools.

References

Huluhulu, N. 2023. Announcement of SA Word of the Year 2023. <https://www.pansalb.org/announcement-of-sa-word-of-the-year-2023/>: [Accessed on: 23 April 2024].

Khumalo, L. 2015. Semi-automatic term extraction for an isiZulu linguistic terms dictionary. *Lexikos* 25: 495-506.

Kennedy, G.D. 1998. *An Introduction to Corpus Linguistics*. London/New York: Longman.

Sinclair, J. 2005. Corpus and Text: Basic Principles. Wynne, M. (Ed.). *Developing Linguistic Corpora: A Guide to Good Practice*. 1-16. Oxford: Oxbow Books. Available online from <http://ahds.ac.uk/linguistic-corpora/> [Accessed 20 October 2005].