

A Corpus-based Critical Discourse Analysis of Gender Sensitivity in isiZulu: Towards an isiZulu Gender Dictionary

Langa Khumalo

ORCID iD: <https://orcid.org/0000-0002-2694-9105>

Rooweither Mabuya

ORCID iD: <https://orcid.org/0000-0001-6658-975>

Abstract

There is paucity of specialized dictionaries for African languages. In this paper we argue for a gender dictionary in isiZulu using a study that evaluates gender sensitivity in isiZulu. This study uses a judicious mixture of Corpus Linguistics (CL) and Critical Discourse Analysis (CDA) in the analysis of gender sensitivity in the Zulu language and culture. Using these two strands, the study critically analyses language and gender issues in isiZulu. The study is framed against the patriarchal background of the Zulu culture whose prejudice against the female gender is expressed through a unique linguistic custom called isiHlonipho (Dowling 1988; Rudwick & Shange 2006; Finlayson 1982). CDA developed from Critical Linguistics as a theory that focuses on discourse as a ‘form of social practice’ (Fairclough & Wodak, 1997:258). CDA explores the relationship between language, power, and society with a critical focus on the crucial role that context plays in discourse. While this framework has been criticised for its lack of objectivity on the part of the analyst, with regards to preferred choice of examples that suit the assumptions of the analyst, and its focus on the fragmented texts rather than full texts, in this study it is used to frame power relations in language use particularly when it refers to gender in isiZulu. To mitigate the evident limitations of CDA this study uses CL to provide sufficient data and context. From the evidence adduced from the corpora used in this study, we motivate for a gender dictionary in isiZulu.

Keywords: Corpora, Corpus Linguistics, CDA, Frequency, Gender, Isi-

Hlonipho, Patriarchy.

1. Introduction

The study uses corpora (singular *corpus*) to evaluate gender sensitivity in isiZulu within the broad framework of Corpus Linguistics (henceforth CL). A corpus is a critical resource used in modern linguistics for carrying out any linguistic investigation. A corpus is ‘a large collection of linguistic data, either written texts or transcription of recorded speech, which can be used as a starting point of description or as a means of verifying hypotheses about a language’ (Crystal 1991:89). A corpus is thus a carefully designed and systematically collected natural language data from a variety of text types and sources following a particular set of principles, which constitutes a sample or dataset that statistically reflects the use of that language and is stored and accessed by means of computers. The size of the corpus and the source from which it is created depends on the intended purpose (Khumalo 2020).

According to Mathye (2003), the fact that gender is a creation of society suggests that the notion of male dominance and female subordination must be viewed from a cultural perspective, as it has no biological basis. Gender plays an important role within the Zulu culture. Language as a carrier of cultural values and biases provide the prism through which one can understand the power and gender dynamics in isiZulu (Wodak 1997).

In traditional patriarchal Zulu culture, women are subordinate to men in terms of power relations. Customary gender roles are expressed through isiHlonipho, which is a unique social and linguistic system that stipulates what constitutes respectful and appropriate language use. IsiHlonipho is a quintessential example of how patriarchy manifests itself through language in the Zulu culture. When viewed through the prisms of CDA, this linguistic practise presents a skewed view of language as an instrument to exercise the power of one gender over the other.

It is evinced from the corpus that isiZulu celebrates bravery, with the most frequent words being *amabutho* (warriors), *imikhonto* (instruments of war) and even human names *Bhekimpi* (Face war/Look after the army/Ready for attack). The most frequent female reference is *indlovukazi* (Mother of the King/Wife of the King). It is evident using the CL that the references to bravery as cited above are associated with maleness. The study also establishes that the use of isiZulu is gendered in that images of power, wealth, precision, success, strength, war, conquering, etc. are all associated with the male gender.

It is further discerned from the study that such values that extoll patriarchy are difficult to shift in a language or culture.

It is further evinced from the frequency list that an argument for an isiZulu gender dictionary that can help explicate the etymology, meaning and context of such a culturally loaded lexicon is motivated. For example, the gendered cultural use of the word *iziko* ‘fireplace’ (referring to women) and *insika* ‘pillar’ (referring to men) can be fully appreciated when the etymology of the two words is fully traced, explicated, and understood.

2. Some Principles for CDA

The CDA framework is useful in analyzing the social and political issues that are constructed and reflected in discourse. Using CDA it can be observed that power relations are negotiated and performed through the way that discourse is formulated. In this framework, discourse both reflects and reproduces social relations. Ideologies are produced and reflected in the use of discourse.

Figure 1.

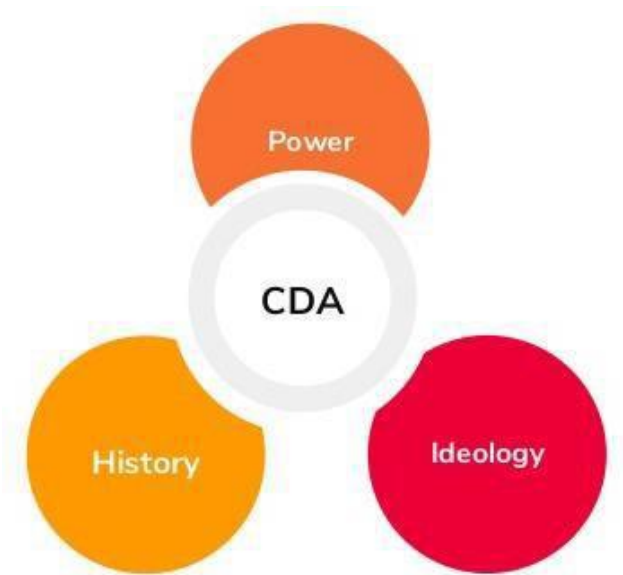


Figure 1 above is a schematic representation of how discourse is an important dataset, whose analysis using CDA provides evidence for relations that have to do with power, history, and ideology. The isiZulu discourse as gleaned from the corpus articulates with the history, ideology, and power relations of the Zulu people. Factored out from the discourse and also represented in the discourse, each of these three dimensions of the discourse captures data that is relevant to the specific dimensions of the CDA. And here, the figure is an illustration of how CDA both represents and frames the history of a people, their ideology and power relations through language, which serves as both a historical and a functioning repository.

Used in the study of Linguistics discourse, contextually, or situation-specifically, CDA therefore explores the relationship between language, power, and society with a critical focus on the crucial role that discourse plays in specific contexts and situations, as these are integrated with specific locations of languages use. So, focused on isiZulu gender sensitivity as we are doing in this article, CDA is therefore used in this study to both frame power, historical and ideological relations in language use particularly when it refers to gender in isiZulu, and also provides dataset information relevant to the specific dimension of analysis. Despite its help and usefulness for analysis, CDA also has its own limitations, because it cannot provide comprehensive data for comprehensive analysis of the dimensions of history, power and ideology. Even so, the limitations are mitigated in the use of CL to provide sufficient data and context, for the CDA.

3. Corpus Linguistics

Corpus Linguistics is variously viewed in the literature as a discipline of linguistics or a branch of linguistics. It is also viewed as a theoretical framework used for linguistics analysis, or alternately as methodological framework that is used as an apparatus to pursue research on any aspect of language, (Kennedy 1998). In this study we use CL as a methodological framework. It is a holistic method that relies on the use of a corpus.

Inherent in the definition of a corpus is the fact that it is a huge purposefully compiled dataset, that is stored digitally by means of computers, and that is queried using efficient and high-speed software programmes. The dataset is made up of naturally occurring language examples. These language examples can be speech data that is recorded and transcribed, or literary data

that is sourced from various genres of writing, that is processed and stored in a computer. The dataset must be representative of the language. This means that it must reflect or mirror the natural language as spoken or written by the community.

The CL method provides in this study authentic examples of isiZulu language use. It further provides the computational affordances that are inherent in CL in analysing the Zulu language and culture. Using the software programme, the analysis is done by analysing the frequency use, the concordance outputs and the collocations in the Zulu language and culture.

The study used of a software tool called WordSmith Tools (version 7) <https://lexically.net/downloads/version7/HTML/index.html>. WordSmith Tools (hence forth WS Tools) is an integrated suite of three main programmes, which include the WordList, Concord, and Keywords. These programmes are used in analysing words and word patterns in any given dataset or corpus. WS Tools was developed by Mike Scott, who had earlier worked with Tim Jones to develop MicroConcord. WS Tools was first released in 1996 and the current version (version 7.0) was released in 2016. The Wordlist programme in WS Tools can be used to produce wordlists or word-cluster lists from a corpus or dataset and render the results in an alphabetic order or by frequency order. It can also calculate word spread across a variety of texts. The Concord programme in WS Tools can give any word or phrase in context – so that one can study its co-text. This is to say that one can see what other words occur in the vicinity of the keyword-in-context. The Keywords programme in WS Tool calculates words which are key in a text, i.e., words that are used much more frequently or much less frequently in each corpus (e.g., the Analysis corpus henceforth the AC) than expected in terms of a general corpus of the language (e.g., the Reference corpus henceforth the RC). This study profited from the enabling tools of this software in analysing data.

4. Patriarchy

Patriarchy refers to the male domination both in public and private spheres. Patriarchy, in its wider definition, means the manifestation and institutionalization of male dominance over women and children in the family and the extension of male dominance over women in society in general. It implies that ‘men hold power in all the important institutions of society’ and that ‘women are deprived of access to such power’ (Sultana 2012:3).

It is evident that patriarchy is a system of social structures and practices in which men dominate, oppress and exploit women. It is a system that is embedded in gender, and the assumption that the male gender is more powerful and dominant over the female gender. This perception or assumption has no evident biological evidence adduced to support it. Using corpus data and the prisms of CDA, the study shows that isiZulu provides sufficient evidence that it is a language that reflects patriarchy.

5. Gender

Gender can be viewed as cultural meanings that are attached to a particular society, its norms, and traditional practices. This view morphs into the way the society conceptualizes the masculine and feminine roles, which in turn influence their personal identities. For example, the view of man, woman, transgender, intersex, etc., among others, in society. This understanding then finds expression in the language of the society or its various discourses. Language becomes a carrier of these views, norms, and practices. It is through studying isiZulu that we conclude that it is a gendered language that is steeped in patriarchy.

According to Mathye (2003), the fact that gender is a creation of society suggests that the notion of male dominance and female subordination must be viewed from a cultural perspective, as it has no biological basis. This also means that this practice may be subject to change as society evolves. This study is therefore crucial in not only establishing that it isiZulu is a gendered language, but also in motivating for the documentation of this phenomenon through a specialized dictionary so that this (historical) knowledge of a people can be stored and documented for both the full meaning of its linguistic expressions and for posterity.

Language is an important vehicle that mirror human assumptions and experiences. We can observe from the foregoing that gender plays an important role within the Zulu culture. Language as a carrier of cultural values and biases provide the prism through which one can understand the power and gender dynamics in isiZulu. We restate our hypothesis that Zulu cultural prejudice against the female gender is expressed through a unique linguistic custom called isiHlonipho. IsiHlonipho is a traditional custom that impels women to respect their male counterparts in a linguistic norm that submits the former to the latter.

6. IsiHlonipho

IsiHlonipho is a linguistic phenomenon that is unique to some South African indigenous languages, in particular the Nguni group. It is widely viewed in isiZulu as a language of respect. It is a linguistic practice that involves the use of a particular vocabulary and manner of speaking that is specific to a particular gender (Makoni 2014). It entails a careful selection of vocabulary by female speakers when referring to male members of their society to reflect respect or politeness. Females have little choice in deciding whether to use isiHlonipho as they command little power in a patriarchal social system (Rudwick & Shange 2006).

It is the presence of this register that motivates the study to dig deeper in the language to see the extent to which isiZulu is generally a language that is gendered. The choice of the corpora used in this study is carefully done to show balance in terms of the historical period, itself a proof to show that patriarchy is a traditional practice that still finds expression in the language.

The hypothesis is that patriarchy is a historical norm, and the choice of early isiZulu dataset was expected to show that it is confined to that past epoch. The contemporary writings that make up the other dataset was expected to show a discourse that represent gender balance or gender neutrality. The analysis of both corpora will show that isiZulu still reflects patriarchy. It will also show etymological information that can best be represented in a specialized dictionary for posterity.

7. The CL Analysis

In this study we used two types of corpora, one Reference Corpus (RC) and the other the Analysis Corpus (AC). The RC was a sample of the IsiZulu National Corpus (INC). This sample had 2.5 million running words. The AC was of two types, one was made up of literature written in the past (early writers of isiZulu literature) and the second AC was made up of contemporary literature. Authors in both epochs were balanced in terms of gender, that is, one male author and one female author. Table 1 below shows the statistics for the Reference and Analysis Corpora (RC & AC).

The ACs were used to compare gender sensitivity in early literary writings to that in contemporary literary writings. This was done to investigate any language shift in terms of gender sensitivity. The RC was used to calculate keyness between AC and RC frequency lists. Keyness is calculated by

‘comparing the frequency of each word in the word list of the text under investigation with the frequency of the same word in the reference word list’ (Scott & Tribble 2006:92). Calculations are done using the Keyword programme of the WS Tools. The output is a list of keywords, or words whose frequencies are higher in the AC than in the RC.

Table 1: RC and AC

File name	File number	Period	Author	Word tokens
Insila kaShaka	AC0001	First isiZulu Novel	J.L. Dube, 1930	19 425
Bafa baphela	AC0002	First isiZulu Novel by female author	J. Gwayi, 1973	22 103
Umdonsiswano	AC0003	Contemporary by male author	C. Nxaba, 2007	49 341
Amandl’ esambane	AC0004	Contemporary by female author	B. Ngcobo, 2009	31 251
			TOTAL AC	122 121
RC	Sample INC (25 million tokens)			2 564 669

Table 2 below shows the wordlist made from the second Analysis Corpus (AC0002). Below the wordlist is the size of corpora (AC0001, AC0002, AC0003 and AC0004). These AC corpora were carefully selected. AC0001 is from the first written isiZulu novel. AC0002 is from the first female isiZulu author. AC0003 and AC0004 are respectively from the male and female contemporary authors. The motivation for this choice was to analyse if the first early writers of isiZulu would depict a discourse that is steeped deep in isiZulu culture, and to observe shifts from such discourse if any in the discourse of the contemporary isiZulu writers. The contemporary isiZulu writers are prominent authors, whose works are widely read, hence their obvious choice.

Table 2: AC0002 Wordlist

N	Word	Freq
18	Izinkomo	58
53	Umuzi	32
56	Abafana	30
61	Izinsizwa	29
66	Umfana	28

File name	File number	Period	Author	Word tokens
Insila kaShaka	AC0001	First isiZulu Novel	J.L. Dube, 1930	19 425
Bafa baphela	AC0002	First isiZulu Novel by female author	J. Gwayi, 1973	22 103
Umdonsiswano	AC0003	Contemporary by male author	C. Nxaba, 2007	49 341
Amandl' esambane	AC0004	Contemporary by female author	B. Ngcobo, 2009	31 251
			TOTAL AC	122 121
RC	Sample INC (25 million tokens)			2 564 669

The most frequent words in the AC0002 with over 22 000 tokens, which is made up of texts from the first novel by a female author in the Zulu language, are typically izinkomo (cattle), umuzi (home), abafana (boys), izinsizwa (young men), and umfana (boy). All these words are in the top 100 most frequent words. These most frequent words clearly show the dominant male gender as adduced from the corpus. The other dominant forms izinkomo (cattle) and umuzi (home) are analysed in 7.1 below using the concordance programme to evaluate the meaning of their high frequency.

7.1 Concordances and Collocates

- 1) Injongo yethu ukuyophanga izinkomo laphaya njengoba wonke
- 2) Mantatisi useziqoqe zonke izinkomo zesizwe; amanye amabutho
- 3) abafana ababafice belusa izinkomo zesizwe sakwaMoloi edlelw
- 4) alelile, bazikhomba zonke izinkomo zesizwe sakwaMoloi, beben
- 5) hweni akhe awashiye elusa izinkomo. Endleleni wathuma enye in
- 6) kotha eyikothayo. Izinkomo zikaMatiwane zase zihamba
- 7) nje indaba yokuba amudle izinkomo uMatiwane. uMatiwane wahl
- 8) dwa inkosi yenu ukungipha izinkomo bese ibuye izilanda ngale
- 9) into esingaba sisayenza izinkomo sezingezami. Nanxa yayis
- 10) amaHlubi eyintshontshela izinkomo zayo yaqala ukubona manje
- 11) honisa kwelamaHlubi ezabo izinkomo. Ngenxa yabo manje ngeke
- 12) isebenzisa ukuze abuyise izinkomo zami azintshontshayo.

The concordances are extracts from the corpus that are extracted using the key-word-in context, izinkomo (cattle). It can be evinced from the concordance lines (1 -12) (in red) and the collocates (in yellow) in 7.1 that the cattle in isiZulu are associated with conquering (line 1), warriors (line 2), boys (line 3), the king (line 8) and stealing (line 10) as summarized in Table 3 below. From the evidence adduced from the corpus, cattle depict dominance, maleness, power and wealth in isiZulu. The association of maleness and various forms of power, and the instruments of power such as wealth, clearly shows the patriarchal domination in the history of the Zulu people.

7.2 AC0003 (Word List 1 – 150)

Table 3.

No	Zulu	English
3	ukuyophanga	conquering
4	amabutho	warriors
5	abafana	boys
16	inkosi	king
18	eyintshontshela	stealing

It is evinced from the AC0003 that isiZulu celebrates bravery. The most frequent words being **N 4 *amabutho*** (warriors), **N 134 *imikhonto*** (instruments of war) and even human names **N 150 *Bhekimpi*** (Face war/Look after the army /Ready for attack). The most frequent female reference is **N 5 *indlovukazi*** (Mother of the King/Wife of the King). It is evident that the meaning of *indlovu-kazi* is again associated with maleness. Images of power, wealth, precision, success, strength, war, conquering, etc. are all associated with the male gender.

8. Motivating the Gender Dictionary

The global focus on gender in the past decade has been interesting, often controversial and steeped in power relations (Ristvedt 2013). The proposed dictionary will draw from the corpus and will evidently cover aspects of African cultural and indigenous knowledge systems (IKS). The dictionary will focus on a lot of assumed knowledge on the role of women by looking at what the data says as shown in Table 4 below. Table 4 shows a list of top 100 most frequent content words in the AC0004 corpus.

Table 4: AC0004 Wordlist 1 – 100

N	Word	Freq
31	Umfazi	58
53	Umuzi	32
77	Iziko	30
94	Umfula	21
100	Insika	14

These words seemed relatively low in the wordlist (except for *umfazi*). The Concordance gave very little context because of the relative smallness of the AC0004. Collocations linked *umfazi* *female*, *iziko* *hearth* and *umfula* *river* to female gender while *umuzi* *home*, was linked to both genders and *insika* *pillar/head* was linked to the male gender.

8.1 Experiment: User Perspective

Using evidence adduced from the corpus we were able to formulate the following definitions for *umfazi* (female), *iziko* (hearth) and *insika* pillar/head), respectively example 1, 2 and 3 below. The headword is in bold, followed by grammatical information (noun and class number), followed by the global definition of the headword (which is drawn from the Oxford Learners' Dictionary), and then followed by the gendered meaning of the headword (which is drawn from the corpus), and finally some etymological reference.

Example 1

umfazi n 2. GLOBAL DEFINITION e.g. an adult female human (Oxford Learners' Dictionary). GENDERED DEFINITION a person who is weak or a pejorative reference to a male person who is perceived to be weak. ETY-

MOLOGICAL INFORMATION e.g. derived from two verbs – fa ‘die’ and –azi ‘know’ = umfazi meaning a person who dies with (a secret) knowledge. Also, full of secrets (= formulated in a positive way: Also, full of knowledge).

The definition in example 1 above associates weakness with the female gender. It provides that the same word umfazi can be used pejoratively when used in reference to the male gender. It also associates negative stereotype such as ‘secret’ with the female gender. It is evident that the framing of the gendered definition derives from an isiZulu corpus, that is steeped in patriarchy.

Although this is not adduced from the corpus, the word umfazi obtains a positive association when it refers to a married woman. It would seem that it gains this positive meaning because the woman has attained a married status, otherwise umfazi loosely refers to a ‘girl or woman’ who has slept with a man outside of wedlock.

Example 2

iziko n 5. **GLOBAL DEFINITION** e.g. the space for a fire in the wall of a room. (Oxford Learners’ Dictionary). **GENDERED DEFINITION** e.g. a person who is the source of love, life, protection and guidance. Often a mother, aunt, or a grandmother. **ETYMOLOGICAL INFORMATION** e.g. derived from the fact that traditionally and symbolically the fireplace or hearth is a social centre of the home; a place associated with warmth, kinship and ancestral spirits. Here food is prepared and shared, stories are told, and knowledge passed from one generation to the next.

The gendered definition presented in example 2 views the female gender as nurturing, and the source of life and love. It also has a symbolic reference to the gendered roles, wherein the general view of the female gender is that it belongs to the kitchen.

Example 3

insika n 5. **GLOBAL DEFINITION** e.g. a large round stone, metal or wooden post that is used to support a bridge, the roof of a building, etc. (Oxford Learners’ Dictionary). **GENDERED DEFINITION** e.g. a strong man who provides strength and support to his family often against all forms of adversity.

ETYMOLOGICAL INFORMATION e.g. derived from the strongest pole that supports the structure of a traditional African hut. EXAMPLE. *Phela kusuke sekukhothame insika nenhloko yomuzi lowo.* (After all, the pillar and head of that family would have died). *IsiZulu Oxford Living Dictionaries.*

This example associates maleness with strength and leadership. The male gender is viewed as the leader of the home, the pillar of the family. It can be argued that looking at these three test examples, the way gender is viewed in isiZulu is based on the patriarchy of the Zulu culture and society. Because culture is evolving and consequently gendered roles evolve, it is expected that this view of maleness and femaleness is evolving. A specialized gender dictionary of isiZulu will provide useful information on the Zulu culture that derives from the literature as evidenced from the corpus.

9. Conclusion

IsiZulu reflects patriarchy as far as this phenomenon is embodied in the gendered isiHlonipho register. The male gender is associated with positive attributes such as power, bravery, wealth, strength, and leadership. The female gender is represented through the prisms of maleness e.g. *indlovukazi* (*Lit. female elephant* = the King's mother/wife) if not generally absent from the entire discourse altogether. Negative attributes such as weakness are associated with the female gender or pejoratively connected to maleness. The only positive association of the female gender is observed in the 'kitchen' (hearth). Women in this setting are viewed as the source of love, life, and the symbol of warmth. The association of female gender with these positive attributes, while welcome is betrayed by the enduring association of the role of women in Zulu society as confined in the kitchen or limited to the home. Thus, an isiZulu gender dictionary is motivated, and it is envisaged that it will preserve the Zulu cultural knowledge for posterity.

References

- Crystal, D. 1991. *A Dictionary of Linguistics and Phonetics*. London: Blackwell Publishers.
- Dowling, T. 1988. *IsiHlonipho Sabafazi: The Xhosa Women's Language of Respect. A Sociolinguistic Exploration*. MA Thesis, University of Cape Town, South Africa.

- Fairclough, N. & R. Wodak 1997. Critical Discourse Analysis. In T.A. van Dijk (ed.): *Introduction to Discourse Analysis*. London: Routledge.
- Finlayson, R. 1982. Hlonipha – The Women's Language of Avoidance among the Xhosa. *South African Journal of African Languages* 1,1: 35 - 60.
<https://doi.org/10.1080/02572117.1982.10586458>
- Kennedy, G.D. 1998. *An Introduction to Corpus Linguistics*. London/ New York: Longman.
- Khumalo, L. 2020. Corpora as Agency in the Intellectualisation of African Languages. In Wolff, H.E. & R.H. Kaschula (eds.): *The Transformative Power of Language*. Cambridge: CUP.
<https://doi.org/10.1017/9781108671088.014>
- Mabuya, R. 2018. *A Corpus Linguistic Analysis of Gender Sensitive Language in isiZulu*. MA Thesis, University of KwaZulu-Natal, South Africa.
- Makoni, B. 2014. Feminizing Linguistic Human Rights: Use of *isihlonipho sabafazi* in the Courtroom and Intra-group Linguistic Differences. *Journal of Multicultural Discourses* 9,1: 27 - 43.
<https://doi.org/10.1080/17447143.2013.806514>
- Mathye, H.R. 2003. The image of women in selected Tsonga novels. MA Thesis, University of South Africa, South Africa.
- Oxford Learners' Dictionary. <https://www.oxfordlearnersdictionaries.com/>
Oxford Living Dictionary, IsiZulu.
- Ristvedt, S.L. 2014. The Evolution of Gender. *JAMA Psychiatry* January 2014, 71, 1: 14 - 15. <https://doi.org/10.1001/jamapsychiatry.2013.3199>
PMid:24227264
- Rudwick, S. & M. Shange 2006. Sociolinguistic Oppression or Expression of 'Zuluness'? 'IsiHlonipho' among isiZulu-speaking Females. *Southern African Linguistics and Applied Languages Studies* 24,4: 473 - 482.
<https://doi.org/10.2989/16073610609486435>
- Scott, M. & C. Tribble 2006. *Textual Patterns: Key Words and Corpus Analysis in Language Education*. Amsterdam: John Benjamins Publishing. <https://doi.org/10.1075/scl.22>
- Sultana, A. 2012. Patriarchy and Women's Subordination: A Theoretical Analysis. *Arts Faculty Journal* 4: 1 - 18.
<https://doi.org/10.3329/afj.v4i0.12929>
- Wodak, R. 1997. *Gender and Discourse*. London: Sage Publications.
<https://doi.org/10.4135/9781446250204>
PMCID:PMC1060568

Towards an isiZulu Gender Dictionary

Langa Khumalo
South African Centre for Digital Language Resources
North-West University
Potchefstroom
South Africa
langa.khumalo@nwu.ac.za

Rooweither Mabaya
South African Centre for Digital Language Resources
North-West University
Potchefstroom
South Africa
Roo.Mabaya@nwu.ac.za