

A CORPUS-BASED INVESTIGATION OF THE ZULU NOMINAL SUFFIX -KAZI: A PRELIMINARY STUDY

**RACHÉLLE GAUTON, GILLES-MAURICE DE SCHRYVER
AND LINKIE MOHLALA**

1 INTRODUCTION

Aarts (2000: 17) states:

“In linguistics, something unprecedented has taken place over the past few decades, something that is perhaps of equal importance for linguistics as, say, the introduction of the electron microscope was for some other disciplines. I am referring of course to the introduction of computer-readable corpora, and even more important, the subsequent explosion in number and size of language corpora that we have witnessed during the last decade. The result has been a mind-boggling increase both in quantity and variety of language data that are now easily accessible for the purposes of language research. Such a vast increase of the availability of actual language data cannot but have a great impact on language research.”

This exciting new development has also not passed South Africa by, where most of the corpus work currently under way centres around the University of Pretoria (UP), and specifically in the department of African Languages. To date, relatively large corpora based on structured samplings of written language have been built for all eleven of the official South African languages, including the nine indigenous Bantu languages. The sizes of these corpora are in constant evolution. For the latest developments, we refer the reader to the Home Page of the Department of African Languages, UP.

Moreover, in a recent sequel of five articles, De Schryver and Prinsloo laid the foundations of a discipline of corpus linguistics for the South African Bantu languages (cf. De Schryver & Prinsloo 2000a, 2000b, 2000c; Prinsloo & De Schryver 2001a, 2001b).

As far as Zulu is concerned, the University of *Pretoria Zulu Corpus* (PZC) is an electronic corpus of 5 million running Zulu words (tokens), organised chronologically and consisting of a number of sub-corpora stratified according to genre. Note that due to a conjunctive writing method, a corpus of 5 million Zulu words is equivalent to an English-language corpus of 6.6 million words (cf. De Schryver & Gauton 2002: 202).

Aarts (2000: 17) indicates that “the time is now ripe for syntactic descriptions that are essentially different from those that have been written so far”. In this paper, we will show how PZC can be queried to achieve exactly this aim in the description of the nominal suffix **-kazi** in Zulu. This suffix is employed to derive feminine and augmentative forms from nouns (as well as, to a lesser extent, from other parts of speech such as relatives and adjectives), and can also express amelioration or derogation, e.g.:

(1) **Feminine (can have positive or negative connotations, depending on the context):**

intando	‘love charm’ + -kazi	>	intandokazi	‘favourite (wife)’
indlovu	‘elephant’ + -kazi	>	indlovukazi	‘queen; she-elephant’
umfelwa	‘widower’ + -kazi	>	umfelokazi / umfelwakazi	‘widow’

Augmentative (can have positive or negative connotations, depending on the context):

indlu	‘house’ + -kazi	>	indlukazi	‘a very big house’
umuthi	‘tree’ + -kazi	>	umuthikazi	‘an outstanding tree’
izinjobo	‘loin-coverings’ + -kazi	>	izinjobokazi	‘ beautiful loin-coverings’
isiziba	‘deep pool’ + -kazi	>	isizibakazi	‘a very deep / dangerous pool’

2 METHODOLOGY FOLLOWED

Through a query of PZC it is possible to make observations regarding the meaning and function of the language feature **-kazi** as reflected in a (written) corpus of actual language usage.

The first step in undertaking this pilot study was to collect all views regarding the use and semantic significations of the suffix **-kazi** from standard (written) sources on the Zulu language, as well as from mother-tongue speakers of the language.

We then assigned a code to each variable, and drew up a matrix reflecting each and every permutation. This did of course lead to a certain amount of over-generation, but our aim was to arrive at conclusions as dictated by the data, and not as determined by any preconceived notions that we as researchers might hold, and/or that might be based on the prevailing views of this grammatical category in standard sources on the language. Subsequently, we queried the 5-million-word Zulu corpus with appropriate corpus query software (in this case *WordSmith Tools*).

The results of such corpus queries can be presented in a multitude of ways by modern corpus query software packages, yet one of the most popular means to look at the proffered output is through the study of KWIC concordance lines. KWIC stands for ‘keyword in context’ and is basically the result of a smart way to automatically cull all instances of a particular word, phrase or morpheme – where each of those can take various wildcards – from an electronic corpus, and to present those strings with co-text on both sides. Together with the searched-for keyword (yet it is clear that it need not be a ‘word’) one can also specify context strings that should (or should not) appear within a specified horizon.

The output, i.e. the concordance lines containing instances of **-kazi**, was then exported to a database. Each and every occurrence was subsequently studied with reference to (a) its

intrinsic meaning (i.e. when taken in isolation), (b) its meaning in context, and then (c) described with reference to each of the following categories:

- Part of speech.
- Vowel ending.
- Quality / value / meaning, i.e.:
 - Negative;
 - Feminine;
 - Augmentative;
 - Feminine + Augmentative (combined significance);
 - Positive.
- Regarding the Feminine, Augmentative, and Feminine + Augmentative categories, an indication was also given as to whether these forms had a Positive, Neutral or Negative connotation.
- Regarding all of the above categories, an indication was given as to which type of noun it had been derived from, i.e.:
 - [+human];
 - [+domestic animal];
 - [+wild animal];
 - other.
- Regarding the Feminine, and Feminine + Augmentative categories, we indicated whether these forms were derived from the masculine or the common forms of the noun types as referred to in the previous point.
- Regarding the Augmentative, and combined Feminine + Augmentative categories, we indicated whether the augmentative signified:
 - added value / importance / intensity;
 - increase in size / quantity;
 - neither of the above.

3 A SUMMARY OF THE MOST IMPORTANT FINDINGS

3.1 TESTING PREVAILING VIEWS AGAINST THE CORPUS

By testing the prevailing views regarding the suffix **-kazi** as found in standard reference sources on the Zulu language against the 5 million word electronic Zulu corpus, we are in a position to arrive at a description of this category that is conditioned not by preconceived notions, introspection or anecdotal data, but that draws upon a vast storehouse of actual language use that fresh observations can be based on in a manner as dictated by the data themselves.

In order to illustrate this process here, we will concentrate only on the use of the suffix **-kazi** with *nouns*. A complete treatment of this category (including its use with adjectives, relatives and adverbs) will be the subject of a forthcoming endeavour.

We found 11857 occurrences of the suffix **-kazi** affixed to nouns in the 5-million-word PZC, which amounts to 89% (11857 / 13368) of all examples containing the nominal suffix **-kazi**.

The corpus study confirms what is implicit to Taljaard & Bosch (1988: 144 *et seq.*) and Doke's (1973⁶: 70 *et seq.*) discussions of this suffix, and is explicitly stated by Wanger (1917: 138) and Van Eeden (1956: 725 & 726), namely that the primary significance of the suffix **-kazi** is the expression of the feminine form, with the augmentative significance as secondary. We found that the primary function of **-kazi** in about 73% (8646 / 11857) of cases is the derivation of the feminine form, whilst in only approximately 12% (1397 / 11857) of cases, **-kazi** is used to derive augmentatives.

As part of this pilot study, we also attempted to ascertain current, spoken mother-tongue usage of this suffix and its relative frequency; for, as stated by Matthews (1925: 1173):

“[...] language is never in the exclusive control of scholars. It does not belong to them alone, as they are often inclined to believe; it belongs to all who have it as a mother-tongue. It is governed not by elected representatives, but by direct democracy, by the people as a whole [...]”

Although our investigation into usage of the suffix **-kazi** in everyday conversation is currently part of an ongoing research project, our preliminary results seem to indicate that mother-tongue speakers of Zulu tend to use the suffix **-kazi** to indicate feminine forms, but make use of nominal and verbal qualifiers such as adjectives, relatives, adverbs, etc. to indicate an augmentative significance – whether signifying an increase in value, importance or intensity, or an increase in size. This preliminary finding is of course firmly in line with what was found in our study of the written corpus, namely that the primary function of the suffix **-kazi** is to signify the feminine form. Furthermore, we are also in the process of trying to ascertain whether there has been any change in usage of the nominal suffix **-kazi** over the course of time. It is expected that these results will lead to various hypotheses regarding the possible diachronic origin of this suffix.

Authorities such as Doke (1973⁶: 73), Van Eeden (1956: 727) and Poulos & Msimang (1998: 110) are in agreement that **-kazi** used as an augmentative, has a relatively restricted application with [+human] and [+animal] nouns, caused by the possibility of confusion with the feminine **-kazi**. When expressing an augmentative meaning with [+human] and [+animal] nouns, other strategies, such as the use of adjectives, are applied. The corpus data bears this out, as it is clear that augmentative forms derived from [+human] and [+animal nouns] account for only 1% (14 / 1397) of cases, as opposed to the overwhelming majority of augmentatives (i.e. 99%, or 1383 / 1397) that are derived from [+inanimate] nouns through the use of the suffix **-kazi**.

The corpus data further support Poulos & Msimang's (1998: 112-113) claims that the feminine form is usually derived from nouns denoting domestic animals by affixing **-kazi** to the generic or common form, whilst with [+human] nouns, the feminine form is derived by suffixing **-kazi** to the masculine form. However, what is not clear from this (and other standard works on the Zulu language) is that the converse also applies, even if only to a limited extent. In our preliminary corpus investigation we have found 256 instances where [+human] feminine nouns are derived from the common or generic form, and 308 instances where the feminine of nouns denoting domestic animals are derived from the masculine form. Compare the following examples in this regard:

(2) **256 instances where [+human] feminine nouns are derived from the common or generic form, e.g.:**

ithishelakazi	‘female teacher’	<	ithishela	‘teacher’
amaXhosakazi	‘Xhosa women’	<	amaXhosa	‘Xhosa people’

308 instances where the feminine of nouns denoting domestic animals are derived from the masculine form, e.g.:

inzimakazi	‘black cow’	<	inzima	‘black ox or bull’
igodlakazi	‘cow with one horn turned downwards’	<	igodla	‘ox with one horn turned downwards’

There are, however, a few claims found in standard sources on the Zulu language that are proven incorrect when tested against the corpus data. Poulos & Msimang (1998: 112) claim that **-kazi** is never used to derive feminine forms from nouns denoting wild animals. In the corpus, however, we found 700 instances of the suffix **-kazi** used to derive the feminine form from nouns referring to wild animals. Example (3) shows a few representative examples of this strategy:

(3) **700 instances of the suffix -kazi used to derive the feminine form from nouns referring to wild animals, e.g.:**

ingwenya	‘crocodile’ + -kazi	>	ingwenyakazi	‘female crocodile’
ukhozi	‘eagle’+ -kazi	>	ukhozikazi	‘female eagle’
impunzi	‘grey duiker (= buck)’ + -kazi	>	impunzikazi	‘grey duiker doe’
inyathi	‘buffalo’ + -kazi	>	inyathikazi	‘buffalo cow’

It must also be noted here that 42 of the 700 examples express *both* a feminine and an augmentative meaning (cf. Van Eeden 1956: 727).

Wanger (1917: 139) states categorically that the feminine suffix **-kazi** does not occur with nouns ending in an **-o**. The corpus data disprove this claim, as there are 206 instances of the feminine **-kazi** suffixed to nouns with a final **-o**. Compare in this regard the following examples:

(4) **206 instances of the feminine -kazi suffixed to nouns with a final -o, e.g.:**

izihlobokazi	‘female relatives’
incokazi	‘red and white speckled cow, skewbald mare’
izimbongolokazi	‘donkey/mule mares’

3.2 OTHER CONCLUSIONS DRAWN FROM THE CORPUS DATA

When studying the corpus, certain aspects of the suffix **-kazi** also come to the fore that have been *under-emphasised*, *inadequately* treated and/or *omitted* in standard works on the Zulu language.

Standard sources on the Zulu language tend to define augmentatives as primarily indicating ‘bigness or greatness’ (cf. Poulos & Msimang 1998: 110). However, it would seem from the corpus data that when **-kazi** is used as an augmentative suffix, it primarily serves to indicate added value, importance or intensity (sometimes in a neutral context, but often in either a positive or a negative context), as opposed to an increase in size. See the following examples in this respect:

(5) **Added value, importance or intensity:**

umpe ‘honey’ > umpekazi ‘particularly tasty honey’

versus

Increase in size:

izimbulunga ‘globes, spheres’ > izimbulungakazi ‘very large globes/spheres’

In all, 67% (939 / 1397) of the augmentatives signify an increase in value, importance or intensity, whilst only 33% (458 / 1397) indicate an increase in size. In the case of nouns that combine a feminine and augmentative meaning, the overwhelming majority of such nouns, namely 96% (51 / 53), express an increase in value, importance or intensity, whilst only 4% (2 / 53) signify an increase in size.

When the occurrences of nouns with the suffix **-kazi** are analysed with a view, firstly to their intrinsic semantic significance, and secondly regarding their significance in context, the following observations can be made.

On the one hand, mother-tongue speakers of Zulu would seem to give feminine nouns derived by way of the suffix **-kazi** an intrinsically positive reading. In context, however, these forms are (more often than not) perceived as displaying a neutral significance, and sometimes a negative significance. Compare the following numerical breakdown in this regard:

(6) **Feminine nouns:**

	Intrinsically		In context	
	#	%	#	%
Positive	6313 / 8641	73%	344 / 8641	4%
Neutral	2323 / 8641	27%	8224 / 8641	95%
Negative	5 / 8641	0.06%	73 / 8641	1%

On the other hand, even though the difference between intrinsic semantic value and use in context is not as significant as is the case with the feminine forms, this type of difference can also be observed with reference to the augmentatives. As can be deduced from

the numerical breakdown in the following table, augmentatives that are perceived as having a negative connotation intrinsically, can have a positive reading, depending on the context.

(7) **Augmentative nouns:**

	Intrinsically		In context	
	#	%	#	%
Positive	11 / 1392	1%	33 / 1392	2%
Neutral	1318 / 1392	95%	1315 / 1392	94%
Negative	63 / 1392	5%	44 / 1392	3%

4 CONCLUSION

In this paper we have shown how a corpus-based approach can lead not only to a more nuanced and accurate description of the Zulu nominal suffix **-kazi**, but can also bring to light aspects of the meaning and function of this suffix that have hitherto been under-emphasised and even overlooked in traditional sources on the Zulu language.

REFERENCES

- Aarts J.** 2000. Towards a new generation of corpus-based English grammars. In Lewandowska-Tomaszczyk B & Melia PJ (eds.). 2000. *PALC'99: Practical Applications in Language Corpora. Papers from the International Conference at the University of Lodz, 15-18 April 1999* (Lodz Studies in Language 1): 17–36. Frankfurt am Main: Peter Lang.
- De Schryver G-M & Gauton R.** 2002. The Zulu locative prefix ku- revisited: a corpus-based approach. *Southern African Linguistics and Applied Language Studies* 20(4): 201-220.
- De Schryver G-M & Prinsloo D.J.** 2000a. The compilation of electronic corpora, with special reference to the African languages. *Southern African Linguistics and Applied Language Studies* 18(1-4): 89–106.
- De Schryver G-M & Prinsloo D.J.** 2000b. Electronic corpora as a basis for the compilation of African-language dictionaries, Part 1: The *macrostructure*. *South African Journal of African Languages* 20(4): 291–309.
- De Schryver G-M & Prinsloo D.J.** 2000c. Electronic corpora as a basis for the compilation of African-language dictionaries, Part 2: The *microstructure*. *South African Journal of African Languages* 20(4): 310–330.
- Doke C.M.** 1973⁶. *Textbook of Zulu grammar*. Cape Town: Longman Southern Africa Limited. *Language Corpora compiled at the University of Pretoria*. Info available from <http://web.up.ac.za/default.asp?ipkCategoryID=1883&sub=1&parentid=482&subid=1866&ipklookid=9>
- Matthews B.** 1925. *Academy Papers*. American Academy of Arts and Letters.
- Poulos G & Msimang C.T.** 1998. *A linguistic analysis of Zulu*. Cape Town: Via Afrika Limited.

- Prinsloo DJ & De Schryver G-M.** 2001a. Corpus applications for the African languages, with special reference to research, teaching, learning and software. *Southern African Linguistics and Applied Language Studies* 19(1-2): 111–131.
- Prinsloo D.J & De Schryver G-M.** 2001b. Monitoring the Stability of a Growing Organic Corpus, with special reference to Sepedi and Xitsonga. *Dictionaries: Journal of The Dictionary Society of North America* 22: 85–129.
- Taljaard P.C. & Bosch S.E.** 1988. *Handbook of isiZulu*. Pretoria: J.L. van Schaik.
- Van Eeden B.I.C.** 1956. *Zoeloe-grammatika*. Stellenbosch: Die Universiteitsuitgewers en -Boekhandelaars.
- Wanger P.W.** 1917. *Konversations-Grammatik der Zulu-Sprache*. Mariannhill: St. Thomas Aquins Druckerei.