Gilles-Maurice de Schryver/Minah Nabirye

TOWARDS A MONITOR CORPUS FOR A BANTU LANGUAGE

A case study of neology detection in Lusoga

Abstract This paper looks at whether, after two decades of corpus building for the Bantu languages, the time is ripe to begin using monitor corpora. As a proof-of-concept, the usefulness of a Lusoga monitor corpus for lexicographic purposes, *in casu* for the detection of neologisms, both in terms of new words and new meanings, is investigated and found useful.

Keywords Monitor corpus; neology detection; new words; new meanings; Bantu; Lusoga

1. Corpus building for the Bantu languages

Corpus building efforts for the Bantu languages remain in their infancy, and not much has changed since the overview published in de Schryver/Prinsloo (2000) – with current corpus sizes typically anywhere between a million and five million tokens. These corpora have mainly been used for dictionary compilation, corpus linguistics, and NLP applications. The last collection of studies in the field of Bantu NLP is already a decade old (De Pauw et al. 2011), and includes studies for all languages from South Africa, as well as Swahili. Recent studies in Bantu corpus linguistics include Dom/de Schryver/Bostoen (2020) for Kisikongo, Kawalya/Bostoen/de Schryver (2021) for Luganda, and Misago/Nshimirimana/Tuyubahe (2021) for Kirundi. Examples of corpus-driven dictionaries compiled for Bantu languages over the past 15 years include de Schryver (2007) for Northern Sotho, de Schryver (2010) for Zulu, and de Schryver/Reynolds (2014) for Xhosa. In all these cases, one or more corpora were built, typically subdivided into a number of sub-corpora reflecting different time periods, genres, and/or topics. The majority of Bantu corpora to date are also 'raw', in that they have not been marked for parts of speech, nor been lemmatised. Also, no project so far has tried to build a 'monitor corpus' for a Bantu language, with which the changing language may be (semi-)automatically tracked (see e.g. Kosem et al. 2021; Kosem 2022). In the current study we attempt exactly that, and apply it to the detection of neologisms in Lusoga, with the aim of improving existing dictionaries for this language.

2. Corpus building for Lusoga

Lusoga is a Great Lakes Bantu language spoken in the Busoga Kingdom, in Eastern Uganda, by about three million people (UBOS 2016, p. 71). Despite a flurry of activity over the past two decades, it may still be classified as a predominantly oral language. During this period, the corpus building effort has been heroically carried forward by a single person (the second author of the present paper), as described in de Schryver/Nabirye (2018). Half a decade ago, the Lusoga corpus stood at a respectable 1.7m tokens (with an oral part of over half a million tokens, 541k more precisely), a corpus mainly used as 'the body of evidence' in writing the first corpus-based grammar of the language (Nabirye 2016). Corpus building continued unabated, and included a special focus on transcriptions of diverse oral data, to reach 3.0m

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

tokens in September 2019 (oral part: 786k; a selection and analysis of which was published in book form: Nabirye 2019).

Within the field of corpus building for the Bantu languages, the Lusoga corpus of 3.0m tokens was considered 'large enough', for it to be able to serve as a base for all future Lusoga studies.[1] Among the tests performed to judge whether or not the Lusoga corpus of 3.0m was also 'stable enough' to act as a reference corpus, stability tests similar to those described by Prinsloo/de Schryver (2001) for the Bantu languages Northern Sotho and Xitsonga were conducted.

Over the past two years, another half a million tokens were collected in addition, bringing the total size of the Lusoga corpus up to 3.5m tokens, nearly a million of them (910k) transcribed material. While it is still a raw corpus, the oral component corresponds to a massive 152 hours of audio recordings; the written component to about 16,000 pages of running text.

3. Towards a monitor corpus for Lusoga

The proof of a pudding is in the eating. It is one thing to judge that a corpus is large and stable enough to be used as a base and reference corpus; it is another entirely to also actually *use* it as such. One valuable use of such a corpus, if it does what it is supposed to do, is to act as a monitor corpus. In their standard textbook, McEnery/Hardie (2012, p. 246) define a 'monitor corpus' as: "A corpus that grows continually, with new texts being added over time so that the dataset continues to represent the most recent state of the language as well as earlier periods." Hanks (2003, p. 53) literally defines a 'dynamic' or so-called 'monitor corpus' in two words: "constantly growing". This may all be good and well and perhaps even feasible for big languages such as English, but for Bantu corpora with their typical modest sizes one can surely not simply keep adding material opportunistically, as one's corpus would lose all its balance and representativeness. As such, given that extreme care is taken to continuously balance out the genres and topics that are being added to the Lusoga corpus, so that it remains representative of both spoken and written Lusoga at all times, some earlier data are sometimes even removed before new material is added (see e.g. de Schryver/Nabirye 2018, § 3.3 vs. § 3.4). In a similar vein, and thus also for reasons of balance and representativeness, 1.6m tokens (of judicial material) have for instance always been kept separate from the main Kirundi corpus (Misago 2018, p. 38), or 2.0m tokens (of religious material) have always been kept separate from the main Luganda corpus (Kawalya 2017, § 5.2 vs. § 5.3 in Chapter 1). In this regard, Kilgarriff's characterisation of how to use monitor corpora for lexicographic purposes is probably more to the point:

> a long-standing vision is the 'monitor corpus', the moving corpus that lets the researcher explore language change objectively (Clear 1988, Janicivic and Walker 1997). The core method is to compare an older 'reference' corpus with an up-to-

[1] In order to put corpus sizes for Great Lakes Bantu languages in context, for the much larger language Kirundi (the national and official language of Burundi, spoken by 8 million people), three scholars contributed to the building of a Kirundi corpus to inform their respective PhDs: Mberamihigo (2014) built and used a Kirundi corpus of 1.9m tokens (oral part: 51k), Nshemezimana (2016) enlarged that to 2.2m tokens (oral part: 196k), and Misago (2018) reached 2.8m tokens (oral part: 418k). For Lusoga's bigger neighbour, Luganda (one of the national languages of Uganda, spoken by 6 million people), Kawalya (2017) built and used a corpus of 4 million tokens for his PhD. Both Kirundi and Luganda have a rich written tradition.

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

> the-minute one to find words which are not already in the dictionary, and which are in the recent corpus but not in the older one. (Kilgarriff 2013, p. 81)[2]

The detection of 'new words' is not the only goal though, as dictionary compilers are also, and sometimes even more so, interested in the detection of new usages, and thus 'new meanings' (cf. Hanks 2002), of existing words:

> Monitor corpora are primarily of importance in lexicographic work [...] They enable lexicographers to trawl a stream of new texts looking for the occurrence of new words or for changing meanings of old words. (McEnery/Wilson 2001, p. 30)

Therefore, and in terms of methodology, we will now compare the additional 0.5m Lusoga material to the earlier 3.0m reference corpus. To do so, we make use of the KeyWords tool from WST (Scott 2019), which calculates the 'outstandingness' of each corpus type. The assumption is that we will be able to detect *new words* which entered the language, as well as *new meanings* for existing words. For the first we assume that we can obtain a limited list of new types in the additional 0.5m that were absent from the 3.0m. For the second we assume that a limited list of 'outstanding' types (specifically types used relatively more frequently over the past two years), will hint at extra usages and thus new meanings. While this exercise may seem trivial, it is not, as what one does not want is long lists of so-called 'new words' that are not new at all but were all simply missing from the 3.0m corpus, and/or 'new meanings' that are not new at all but were all simply not used in the 3.0m corpus. That a certain percentage was truly missing or not used is acceptable (no corpus, no matter its size, contains all the words in all its uses for any given language), but to usefully act as a monitor corpus, a useful percentage must be 'new' words and usages, or thus 'neologisms'. If neologisms are indeed detected in this way, lexicographers may also act upon those. That said, if this exercise is successful – in the sense that it results in meaningful data that can be acted upon by dictionary compilers – we can then consider the 3.5m corpus as the new reference and thus new monitor corpus.

4. The semi-automatic identification of neologisms in Lusoga

4.1 New words

The default settings of WST's KeyWords were used and, fair enough, a limited number of 55 keywords occurring in at least two of the new texts was found that had not been seen in the 3.0m corpus. An analysis of the categories these 55 'new words' belong to is shown in Figure 1.

One of the 'new words', unsurprisingly, is ***COVID***, a clear neologism. (***Corona*** was also picked up, but because there was already a single mention of it – as the "Corona Hospital" (in California) – it was marked as outstanding; see § 4.2.) As with every non-English monitor corpus, the expectation was that a good number of 'new words' would be proper names

[2] The second reference should be to Janicijevic/Walker (1997); and the title of Clear's (1988) paper is "Trawling the language: Monitor corpora" rather than the misquoted "The Monitor Corpus". Both of these errors are unfortunately found all over the metalexicographic literature. More upbeat: As with so much in our field, the term 'monitor corpus' was coined by John Sinclair, in 1982, or thus four decades ago already (Clear 1988, p. 383).

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

and (given that English is the only official language of Uganda) also plain English words. This is borne out; these categories make up 16% and 4% respectively, so 20% in all. Proper names and plain English words are not normally items that warrant inclusion in a Lusoga dictionary.

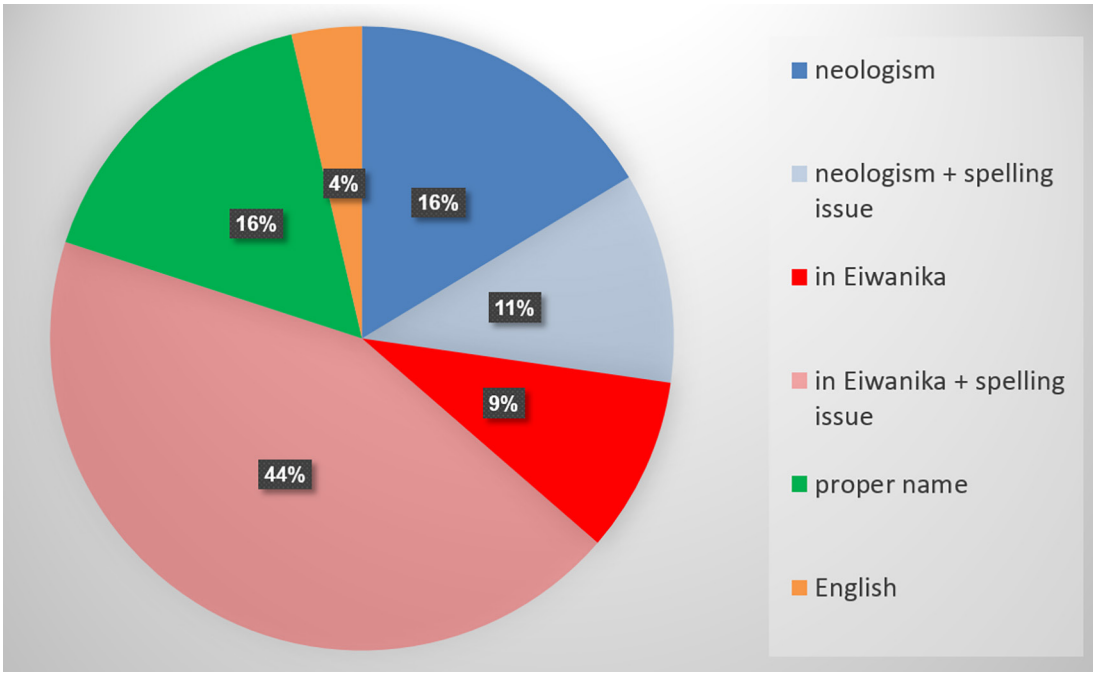

Fig. 1: Analysis of the 'new words' that entered Lusoga between September 2019 and January 2022

In order to analyse the data, it is important to note that no common (standard) orthography has yet been adopted by all those who write in Lusoga, so a tool like WST will also pick up (and give too much weight to) spelling variations – see 'spelling issue' in Figure 1. Had a dictionary been compiled based on the 3.0m corpus, the remaining 80% of the keywords from Figure 1, would all be candidates for inclusion in an update to the dictionary. While such a corpus-based dictionary does not exist, a *non-corpus-based* monolingual dictionary has been compiled, namely the *Eiwanika ly'Olusoga* (Nabirye 2009), online since 2012 at https://menhapublishers.com/dictionary/.

The quest for neologisms may then be rephrased as a quest for candidates to update that dictionary. Astonishingly, as many as 53% of the 'new words' from Figure 1 were already included in the *Eiwanika*, so they are not new words at all; just 27% are. The latter include new loanwords for ***omusaseredooti*** 'priest' (< Latin *sacerdos* 'priest'), ***mwepisikoopi*** 'bishop' (< Latin *episcopus* 'bishop'), and ***ukarisitia*** 'Eucharist' (< Greek *Eucharist* 'gratitude'), but also concepts that can only be 'derived', using language-internal processes, from other words already in the *Eiwanika*, and which are thus debatable neologisms, such as ***obukurisitu*** 'Christianity' (***Omukristo*** 'Christian' is in the *Eiwanika*), ***omuyumo*** 'entertainer' (***ekinhumo*** 'party' is in), or the reduplicated form ***mutoto*** 'youngish' (***-to*** 'young' is in). Conversely, others are clearly true neologisms: ***akanhomero*** 'a small pejorative place' (< ***okunhooma*** 'despise') or ***ekizezengere*** 'shadow' (the personification of ***ekinzenze*** 'a shadow').

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Regarding the first three religious terms here (for 'priest', 'bishop', and 'Eucharist') one may wonder why we label them neologisms, as surely terms for those concepts were already in the language. Suffice it to say that competing religious groups devised their own terms in Lusoga, and that with the recent publication and now inclusion in the Lusoga corpus of Roman Catholic material, these 'new' terms (for old concepts) have now also officially entered the Lusoga language.[3]

4.2 New meanings

In addition to the 55 'new words', WST also lists 1,251 'outstanding words': 815 'positive keywords' (= words that are relatively more frequent in the new 0.5m material compared to the monitor corpus of 3.0m), and 436 'negative keywords' (= words that are relatively less frequent in the new material compared to the monitor corpus, and may thus be 'disappearing from the language'). Of the positive keywords, 466 occur in at least two of the new 0.5m corpus files, while 349 occur in just one of the new files. For the purposes of the present paper, we will only look at the top 100 positive keywords that occur in at least two new texts. An analysis of the categories these 'top 100' belong to is shown in Figure 2.

In Figure 2, the proportion of proper names has slightly grown compared to Figure 1 (to 20%), while that of English has gone down (to 2%). A notable proper name that is now far more outstanding is that of ***Gabula***, the title of the current Busoga King. In terms of candidate new meanings, as many as 61% turn out to have been properly covered in the *Eiwanika*, with their various meanings; yet 17% have not. Some of these 17% indicate that a number of function words which are the result of grammatical constructions had better been lemmatised in the *Eiwanika*, such as the connectives (construction = pronominal prefix + ***-a***), and that some combinations also warrant lemma-sign status, such as ***-liwo*** 'be present' (< ***-li*** 'to be' + ***wo*** (locative)), or ***me ni*** 'and then' (< ***me*** 'and then' + ***ni*** (focus)). These, of course, are neither new words nor new uses; yet the software has (correctly!) picked them out as candidate entries. So here the use of a monitor corpus for Lusoga has not detected new meanings, but forces lexicographers to face the facts; and the fact is that more grammar needs to be entered into the central lemma-sign list of a dictionary.

[3] The work concerned is the Roman missal (Gonza 2018); which despite being dated 2018 only became available in late 2019, whereupon it was scanned, OCRed, and heavily processed (by the first author of this paper, to take out all the English parts) before it was added to the corpus. The Protestant Bible (BSU 2014) was already in the corpus.

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

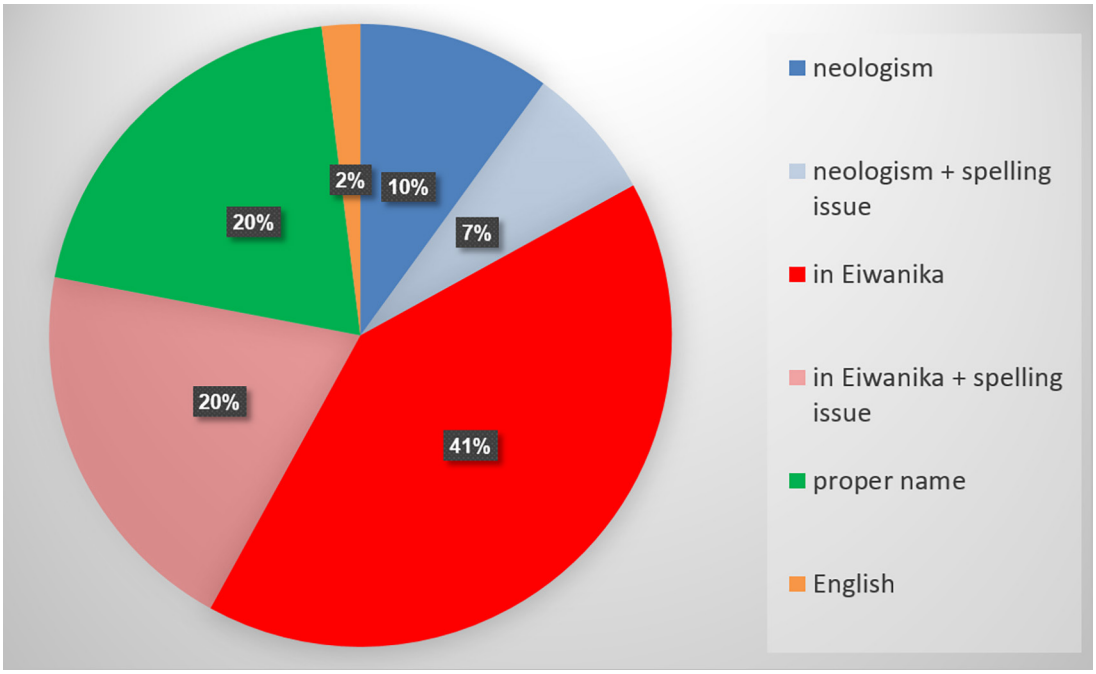

Fig. 2: Analysis of the top 100 'outstanding words' (in at least two corpus files) when comparing Lusoga between September 2019 and January 2022

Full words not lemmatised in the *Eiwanika* include ***lebe*** 'so and so', as well as the interjection ***eee***. The specific but non-descript meaning 'so and so' may be considered a near-neologism; it was hardly there before but now entered the language 'in force'. Similarly for the unspecific interjection *eee*, while not lemmatised in the *Eiwanika*, it *was* used once in a single example (under the lemma ***(a)keewuunia***).

An interesting language change is ***eisakaramentu*** 'sacrament': ***saakalamentu*** was lemmatised in the *Eiwanika*, but the monitor corpus now indicates that the form with a noun class prefix has become far more acceptable than it used to be.

The remainder are all clear cases of neologisms, as these are words that acquired new and very specific meanings. These include: ***ebyeghongo*** 'things used to pray; gifts' (deverbative < ***okuwonga*** 'to give offerings in church'), ***amaingira*** 'the process of entering' (deverbative < ***okwingila*** 'to enter'), ***ekitaloodheka*** 'that which is difficult to relay' (deverbative < cl. 7 noun prefix + negative marker ***-ta*** + ***okuloodha*** 'to relay' + stative extension), ***olugololiro*** 'in a straight manner' (deverbative < ***okugolola*** 'to make straight'), and ***kituufu*** 'it is true' (adjective < ***obutuufu*** 'truthfulness').

5. Discussion and conclusion

As Kilgarriff (2013, p. 82) correctly pointed out: "The nature of the task is that the automatic process creates a list of candidates, and a lexicographer then goes through them to sort the wheat from the chaff. There is always far more chaff than wheat." In terms of 'new words', adding half a million Lusoga tokens to a corpus of 3 million tokens, revealed just 55 items, so having to sort the wheat (which turned out to be 27%) from the chaff manually for such a small amount is more than doable. In terms of 'new meanings', we presented an analysis of the top 100 outstanding words only, where we saw that the wheat was less

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

forthcoming (17%). The full details of going from raw data to analysis may be found in Addenda 1 and 2.[4]

While Kilgarriff does not give us an indication of an acceptable ratio of wheat to chaff, apart from informing us that it is inherently low, we feel that the exercise for Lusoga was worthwhile, as we did pinpoint enough useful material to update the *Eiwanika*. As a result, we are confident that the dawn of monitor corpora for the Bantu languages has arrived.

However, upon also considering recall and precision when going down the list of potential new meanings [to be presented during the actual talk only, as space constraints do not allow for a full description here], we are dealing with a case of diminishing returns: The recall does indeed go up, but at an increasingly punishing precision. Another bottleneck, especially with hopes of automating the process in future, revolves around the various spellings used among the Basoga community; but this is a language-specific problem, not a Bantu-wide one.

References

BSU. 2014. Baibuli. Ekibono kya Katonda. Omuli n'ebitabo ebyetebwa deuterokanoniko/apokurifa [Bible. The word of God, which also has the books known as Deuteronomy/Apocrypha]. Kampala.

Clear, J. (1988): Trawling the language: monitor corpora. In: Snell-Hornby, M. (ed.): ZuriLEX '86 Proceedings: Papers read at the EURALEX International Congress. Tübingen, pp. 383–389.

De Pauw, G./de Schryver, G.-M./Pretorius, L./Levin, L. (2011): Introduction to the special issue on African Language Technology. In: Language Resources and Evaluation 45 (3), pp. 263–269.

de Schryver, G.-M. (2007): Oxford bilingual school dictionary: Northern Sotho and English / Pukuntšu ya Polelopedi ya Sekolo: Sesotho sa Leboa le Seisimane. E gatišitšwe ke Oxford. Cape Town.

de Schryver, G.-M. (2010): Oxford bilingual school dictionary: Zulu and English / Isichazamazwi Sesikole Esinezilimi Ezimbili: IsiZulu NesiNgisi, Esishicilelwe abakwa-Oxford. Cape Town.

de Schryver, G.-M. (2020): Linguistics terminology and neologisms in Swahili: rules vs. practice. In: Dictionaries: Journal of The Dictionary Society of North America 41 (1), pp. 83–104 + 14 pages of supplementary material online.

de Schryver, G.-M./Nabirye, M. (2018): Corpus-driven Bantu lexicography, Part 1: Organic corpus building for Lusoga. In: Lexikos 28, pp. 32–78.

de Schryver, G.-M./Prinsloo, D. J. (2000): The compilation of electronic corpora, with special reference to the African languages. In: Southern African Linguistics and Applied Language Studies 18 (1–4), pp. 89–106.

de Schryver, G.-M./Reynolds, M. (2014): Oxford bilingual school dictionary: IsiXhosa and English / Oxford isiXhosa-isiNgesi English-isiXhosa Isichazi-magama Sesikolo. Cape Town.

Diki-Kidiri, M. (ed.) (2008): Le vocabulaire scientifique dans les langues africaines. Pour une approche culturelle de la terminologie. (= Collection Dictionnaires et Langues). Paris.

[4] Note that it has not been the purpose of this paper to also analyse the various linguistic strategies used to create neologisms in Lusoga, even though some of the evidence for this may be deduced from the Addenda. For a recent Bantu example in this domain, see de Schryver (2020) who deals with term creation processes in Swahili. In Africa more generally, many scholars have worked on this as well; see for instance the edited collection by Diki-Kidiri (2008) for case studies in West and Central Africa.

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Dom, S./de Schryver, G.-M./Bostoen, K. (2020): Kisikongo (Bantu, H16a) present-future isomorphism: a diachronic conspiracy between semantics and phonology. In: Journal of Historical Linguistics 10 (2), pp. 251–288.

Gonza, R. K. (2018): Misaale mu Lusoga. Jinja.

Hanks, P. (2002): Mapping meaning onto use. In: Corréard, M.-H. (ed.): Lexicography and natural language processing. A festschrift in honour of B.T.S. Atkins. Euralex, pp. 156–198.

Hanks, P. (2003): Lexicography. In: Mitkov, R. (ed.): The Oxford handbook of computational linguistics. Oxford, pp. 48–69.

Janicijevic, T./Walker, D. (1997): NeoloSearch: Automatic detection of neologisms in French Internet documents. In: Proceedings of the 1997 Joint International Conference of the Association for Computers and the Humanities and the Association for Literary & Linguistic Computting. Kingston, Ontario, pp. 93–94.

Kawalya, Deo. 2017. A corpus-driven study of the expression of modality in Luganda (Bantu, JE15). PhD dissertation. Ghent.

Kawalya, D./Bostoen, K./de Schryver, G.-M. (2021): A diachronic corpus-driven study of the expression of possibility in Luganda (Bantu, JE15). In: International Journal of Corpus Linguistics 26 (3), pp. 336–369.

Kilgarriff, A. (2013): Using corpora as data sources for dictionaries. In: Jackson, H. (ed.): The Bloomsbury companion to lexicography. London, pp. 77–96.

Kosem, I. (2022): Trendi – a monitor corpus of Slovene. In: Klosa-Kückelhaus, A./Engelberg, St./Möhrs, Ch./Storjohann, P. (eds): Proceedings of the XX EURALEX International Congress: Dictionaries and Society. Mannheim.

Kosem, I./Krek, S./Gantar, P./Holdt, Š. A. /Čibej, J. (2021): Language monitor: tracking the use of words in contemporary Slovene. In: Kosem, I./Cukr, M./Jakubíček, M. /Kallas, J./Krek, S./Tiberius, C. (eds): Electronic Lexicography in the 21st Century (eLex 2021): Post-Editing Lexicography. Proceedings. Brno, pp. 514–528.

Mberamihigo, F. (2014): L'expression de la modalité en kirundi: exploitation d'un corpus électronique. PhD dissertation. Brussels/Ghent.

McEnery, T./Hardie, A. (2012): Corpus linguistics: method, theory and practice (= Cambridge Textbooks in Linguistics). Cambridge.

McEnery, T./Wilson, A. (2001): Corpus linguistics: an introduction. 2nd Edition. (= Edinburgh Textbooks in Empirical Linguistics). Edinburgh.

Misago, M.-J. (2018): Les verbes de mouvement et l'expression du lieu en kirundi (bantou, JD62): une étude linguistique basée sur un corpus. PhD dissertation. Ghent.

Misago, M.-J./Nshimirimana, E./Tuyubahe, P. (2021): Usages grammaticaux du verbe -guma 'rester' en kirundi (JD62): Une étude linguistique basée sur un corpus. In: Language in Africa 2 (1), pp. 3–40.

Nabirye, M. (2009): Eiwanika ly'Olusoga. Eiwanika ly'aboogezi b'Olusoga n'abo abenda okwega Olusoga [A dictionary of Lusoga. For speakers of Lusoga, and for those who would like to learn Lusoga] (= Linguistics Series 1). Kampala.

Nabirye, M. (2016): A corpus-based grammar of Lusoga. PhD dissertation. Ghent.

Nabirye, M. (2019): Owayanga: Empayo Dhimala Dhaavaamu Olufumo [Speak regularly: conversations have a tendency to become legends]. (= Linguistics Series 5). Kampala.

Nshemezimana, E. (2016): Morphosyntaxe et structure informationnelle en kirundi: Focus et stratégies de focalisation. PhD dissertation. Ghent.

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Prinsloo, D. J./de Schryver, G.-M. (2001): Monitoring the stability of a growing organic corpus, with special reference to Sepedi and Xitsonga. In: Dictionaries: Journal of The Dictionary Society of North America 22, pp. 85–129.

Scott, M. (2019): WordSmith Tools, version 7. http://www.lexically.net/wordsmith/.

UBOS (2016): The National Population and Housing Census 2014 – Main Report. Kampala.

Contact information

Gilles-Maurice de Schryver
BantUGent – UGent Centre for Bantu Studies, Ghent University
& Department of African Languages, University of Pretoria
gillesmaurice.deschryver@UGent.be

Minah Nabirye
BantUGent – UGent Centre for Bantu Studies, Ghent University
minah.nabirye@UGent.be

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Addendum 1: Raw data for 'new words' that entered Lusoga between September 2019 and January 2022

Keyword	Translation	Spelling issue?	Other spelling	Eiwan-ika?	Neolo-gism?	Notes	Freq.	Freq. %	Texts	RC freq.	RC %	BIC	Log-likelihood	Log-ratio	Probability
kurisitu	proper name	✓	Kristu	✓			2,126	0.3933	2	0	0.000	8,100.99	8,116.10	141.01	0.00000000000000000002
jamiili	proper name						142	0.0263	4	0	0.000	526.98	542.09	137.11	0.00000000000000006099
omusaseredooti	priest				✓	< Latin **sacerdos** 'priest'	137	0.0253	4	0	0.000	507.89	523.00	137.05	0.00000000000000006842
abakurisitu	Christians	✓	Abakristu	✓			90	0.0166	3	0	0.000	328.47	343.58	136.45	0.00000000000000026963
kaliisi	proper name						60	0.0111	2	0	0.000	213.94	229.05	135.86	0.00000000000000107804
musaseredooti	priest				✓	< Latin **sacerdos** 'priest'	52	0.0096	3	0	0.000	183.40	198.51	135.66	0.00000000000000179678
mbotana	proper name						52	0.0096	2	0	0.000	183.40	198.51	135.66	0.00000000000000179678
oketcho	proper name						50	0.0092	4	0	0.000	175.77	190.88	135.60	0.00000000000000207210
ekiir	abbr. for ekiiribwamu 'response'				✓	**ekiiribwamu** 'response'	50	0.0092	2	0	0.000	175.77	190.88	135.60	0.00000000000000207210
abakuristu	Christians	✓	Abakristu	✓			48	0.0089	2	0	0.000	168.13	183.24	135.54	0.00000000000000240667
mwepisikoopi	bishop				✓	< Latin **episcopus** 'bishop'	45	0.0083	2	0	0.000	156.68	171.79	135.45	0.00000000000000305803
omukwampa	the direction of				✓		44	0.0081	2	0	0.000	152.86	167.97	135.42	0.00000000000000332697
ukarisitia	Eucharist				✓	< Greek **Eucharist** 'gratitude'	35	0.0065	2	0	0.000	118.50	133.61	135.09	0.00000000000000809336
eikukubira	crowd / mob	✓	eikuukuubila	Y			29	0.0054	2	0	0.000	95.60	110.71	134.81	0.00000000000001772373
kategere	proper name						28	0.0052	2	0	0.000	91.78	106.89	134.76	0.00000000000002065687
kukusengera	to praise you	✓	kukusengela	✓			22	0.0041	2	0	0.000	68.88	83.99	134.42	0.00000000000006405400
ainjira	you enter	✓	aingila	✓			20	0.0037	3	0	0.000	61.24	76.35	134.28	0.00000000000010531816
mwoyogwe	his soul	✓	mwoyo gwe	✓			18	0.0033	3	0	0.000	53.61	68.72	134.13	0.00000000000019120947
bukobelana	proper name						18	0.0033	3	0	0.000	53.61	68.72	134.13	0.00000000000019120947
yaacuukira	he changed from	✓	yakyukila	✓			17	0.0031	3	0	0.000	49.79	64.90	134.04	0.00000000000027119964
fridge	English						16	0.0030	2	0	0.000	45.97	61.08	133.96	0.00000000000040281970
kirimungo	proper name						15	0.0028	4	0	0.000	42.15	57.26	133.86	0.00000000000063545448
kuntwiko	on top	✓	ku ntwiko	✓			15	0.0028	2	0	0.000	42.15	57.26	133.86	0.00000000000063545448

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Keyword	Translation	Spelling issue?	Other spelling	Eiwan-ika?	Neolo-gism?	Notes	Freq.	Freq. %	Texts	RC freq.	RC %	BIC	Log-likelihood	Log-ratio	Probability
mwikuukuubira	in a crowd / mob	✓	mw'ikuukuubi-la	✓			15	0.0028	2	0	0.000	42.15	57.26	133.86	0.00000000000063545448
kwigulu	on heaven	✓	ku igulu	✓			14	0.0026	2	0	0.000	38.34	53.45	133.76	0.00000000000108799153
asikaali	guard			✓			14	0.0026	2	0	0.000	38.34	53.45	133.76	0.00000000000108799153
covid	COVID				✓	COVID & COVID-19	13	0.0024	6	0	0.000	34.52	49.63	133.66	0.00000000000209591906
omuyumo	entertainer				✓	**ekinhumo** 'party' is in; but not the entertainer	13	0.0024	5	0	0.000	34.52	49.63	133.66	0.00000000000209591906
omuminsani	missionary			✓			13	0.0024	2	0	0.000	34.52	49.63	133.66	0.00000000000209591906
obukurisitu	Christianity	✓	Obukristu		✓	**Omukristo** 'Christian' is in; but not cl. 14	12	0.0022	3	0	0.000	30.70	45.81	133.54	0.00000000000485604720
bwoisa	as you breathe / as you say	✓	bw'oisa	✓			12	0.0022	2	0	0.000	30.70	45.81	133.54	0.00000000000485604720
leonard	proper name						12	0.0022	2	0	0.000	30.70	45.81	133.54	0.00000000000485604720
okugwana	to deserve			✓			12	0.0022	2	0	0.000	30.70	45.81	133.54	0.00000000000485604720
akanhomero	a small pejorative place	✓	akanhoomelo		✓	**okunhooma** 'despise' is in; but not the deverbative	12	0.0022	2	0	0.000	30.70	45.81	133.54	0.00000000000485604720
omubirigwe	his body	✓	omubili gwe	✓			11	0.0020	3	0	0.000	26.88	41.99	133.42	0.00000000001566624434
acuuka	he changes	✓	akyuka	✓			11	0.0020	3	0	0.000	26.88	41.99	133.42	0.00000000001566624434
buliribwe	his bed	✓	bulili bwe	✓			11	0.0020	2	0	0.000	26.88	41.99	133.42	0.00000000001566624434
abamwaguleku	so that he sprinkles on you				✓	**okumwaga** 'to sprinkle' is in; but not the reversive + subjunctive + loc. form	10	0.0018	2	0	0.000	23.07	38.18	133.28	0.00000000010989710877
filename	English						9	0.0017	9	0	0.000	19.25	34.36	133.13	0.00000001660683479621
bakurisitu	Christians	✓	Bakristu	✓			9	0.0017	3	0	0.000	19.25	34.36	133.13	0.00000001660683479621
bwanaaba	if he will be	✓	bw'anaaba	✓			9	0.0017	3	0	0.000	19.25	34.36	133.13	0.00000001660683479621
simwogerere	proper name						9	0.0017	3	0	0.000	19.25	34.36	133.13	0.00000001660683479621
mwikonero	home	✓	mw'ikonelo		✓	new meaning (was: chair; water container > home > chair; whomever owns the water container owns the home, thus chair > cf. one is not supposed to sit in the parent's chair)	9	0.0017	2	0	0.000	19.25	34.36	133.13	0.00000001660683479621
yaacuuka	he is changing	✓	yaakyuka	✓			9	0.0017	2	0	0.000	19.25	34.36	133.13	0.00000001660683479621
endhuuye	his house	✓	endhu ye	✓			9	0.0017	2	0	0.000	19.25	34.36	133.13	0.00000001660683479621
ebiraalagho	those which remain; those which are maintained	✓	ebilaalawo		✓	**okulaala** 'be calm' is in; -wo = loc.	9	0.0017	2	0	0.000	19.25	34.36	133.13	0.00000001660683479621

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Keyword	Translation	Spelling issue?	Other spelling	Eiwan-ika?	Neolo-gism?	Notes	Freq.	Freq. %	Texts	RC freq.	RC %	BIC	Log-likelihood	Log-ratio	Probability
ekizezengere	shadow	✓	ekinzenzegele		✓	**ekinzenze** 'a shadow' is in Eiwanika; but new personification of 'shadow' is not	9	0.0017	2	0	0.000	19.25	34.36	133.13	0.000000001660683479621
ekiyumu	party	✓	ekinhumo	✓			8	0.0015	8	0	0.000	15.43	30.54	132.96	0.00000029774795962112
bainjira	they enter	✓	baingila	✓			8	0.0015	4	0	0.000	15.43	30.54	132.96	0.00000029774795962112
kyamwidhira	it came to him	✓	kyamwidhila	✓			8	0.0015	3	0	0.000	15.43	30.54	132.96	0.00000029774795962112
twamubaamu	we were in it			✓		**-ba**	8	0.0015	3	0	0.000	15.43	30.54	132.96	0.00000029774795962112
okucaala	to visit	✓	okukyala	✓			8	0.0015	3	0	0.000	15.43	30.54	132.96	0.00000029774795962112
mutoto	youngish	✓	mutooto		✓	**-to** 'young' is in; but not the reduplicated form	8	0.0015	2	0	0.000	15.43	30.54	132.96	0.00000029774795962112
kabbuli	grave (Islam)			✓			8	0.0015	2	0	0.000	15.43	30.54	132.96	0.00000029774795962112
bweyinza	as it can	✓	bw'eyinza	✓			8	0.0015	2	0	0.000	15.43	30.54	132.96	0.00000029774795962112

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Addendum 2: Raw data for the top 100 'outstanding words' (in at least two corpus files) when comparing Lusoga between September 2019 and January 2022

Keyword	Translation	Spelling issue?	Other spelling	*Eiwan-ika*?	Neolo-gism?	Notes	Freq.	Freq. %	Texts	RC freq.	RC %	BIC	Log-likelihood	Log-ratio	Probability
amiina	Amen			✓			1,818	0.3363	9	538	0.017	4,566.20	4,581.31	4.28	0.00000000000000000009
ebyanda	century			✓			1,989	0.3680	6	1,506	0.048	3,283.21	3,298.31	2.92	0.00000000000000000023
ighe	you	✓	iwe	✓			1,360	0.2516	17	523	0.017	3,119.60	3,134.71	3.90	0.00000000000000000027
tubita	we pass			✓			1,447	0.2677	9	892	0.029	2,685.58	2,700.69	3.22	0.00000000000000000042
tukweghembye	we have beseached you	✓	tukwewembye	✓			1,035	0.1915	2	321	0.010	2,554.86	2,569.97	4.21	0.00000000000000000049
tukweghemba	we beseach you	✓	tukwewemba	✓			1,394	0.2579	3	925	0.030	2,484.12	2,499.23	3.11	0.00000000000000000053
mutukuvu	he is holy			✓			1,222	0.2261	4	704	0.023	2,346.92	2,362.03	3.32	0.00000000000000000063
alamula	he passes judgement			✓			700	0.1295	3	62	0.002	2,247.16	2,262.27	6.02	0.00000000000000000072
aligho	he is present	✓	aliwo		✓	**-li** 'to be' is in; but not the combined form	715	0.1323	9	95	0.003	2,159.33	2,174.44	5.43	0.00000000000000000082
waife	ours			✓			1,926	0.3563	26	2,458	0.079	2,113.59	2,128.70	2.17	0.00000000000000000087
nabirye	proper name						2,827	0.5230	23	5,068	0.163	2,103.82	2,118.93	1.68	0.00000000000000000088
okusembera	to come close; to receive communion	✓	okusembela	✓			831	0.1537	3	493	0.016	1,567.30	1,582.40	3.28	0.00000000000000000215
lw	connective 'of'				✓	**-a** 'of' not in	1,438	0.2660	33	2,089	0.067	1,376.38	1,391.49	1.98	0.00000000000000000320
olwembo	song	✓	olwemba	✓			920	0.1702	9	825	0.027	1,347.91	1,363.02	2.68	0.00000000000000000341
omutuukirivu	the one who is holy	✓	omutuukilivu	✓			575	0.1064	3	216	0.007	1,321.79	1,336.90	3.93	0.00000000000000000362
nantambi	proper name						442	0.0818	5	96	0.003	1,198.39	1,213.50	4.73	0.00000000000000000488
mwoyo	soul			✓			1,097	0.2029	15	1,499	0.048	1,117.53	1,132.64	2.07	0.00000000000000000603
okughonga	to pray	✓	okuwonga	✓			575	0.1064	3	323	0.010	1,110.43	1,125.54	3.35	0.00000000000000000615
kaluuba	proper name						345	0.0638	2	30	0.001	1,102.49	1,117.60	6.05	0.00000000000000000629
wo	locative 16			✓			1,582	0.2927	22	3,000	0.097	1,081.22	1,096.33	1.60	0.00000000000000000667
yezu	proper name		Yesu / Yezu	✓			727	0.1345	3	645	0.021	1,070.17	1,085.28	2.69	0.00000000000000000689

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Keyword	Translation	Spelling issue?	Other spelling	*Eiwanika*?	Neologism?	Notes	Freq.	Freq. %	Texts	RC freq.	RC %	BIC	Log-likelihood	Log-ratio	Probability
ebyeghongo	things used to pray; gifts	✓	ebyewongo		✓	**okuwonga** 'to give offerings in church' is in; but not the deverbative	630	0.1165	3	480	0.015	1,025.55	1,040.66	2.91	0.00000000000000000784
abailamu	those who respond			✓			275	0.0509	5	2	0.000	1,011.65	1,026.76	9.63	0.00000000000000000818
ndhote	proper name						284	0.0525	7	16	0.001	949.28	964.39	6.67	0.00000000000000000993
eighanga	country; tribe	✓	eiwanga	✓			651	0.1204	11	598	0.019	932.81	947.92	2.64	0.00000000000000001048
kadaaga	proper name						303	0.0561	5	46	0.001	884.29	899.40	5.24	0.00000000000000001234
kasaata	proper name						248	0.0459	9	9	0.000	856.52	871.62	7.31	0.00000000000000001361
mata	proper name						355	0.0657	14	118	0.004	846.57	861.68	4.11	0.00000000000000001411
omughanzi	the one who is able	✓	omuwanzi	✓			443	0.0820	3	265	0.009	824.85	839.96	3.26	0.00000000000000001528
lebe	so and so				✓	MISSED	334	0.0618	7	104	0.003	813.18	828.29	4.21	0.00000000000000001596
aghalala	together; collectively	✓	awalala	✓			792	0.1465	13	1,076	0.035	807.48	822.59	2.08	0.00000000000000001631
baagalana	proper name						232	0.0429	2	8	0.000	802.98	818.09	7.38	0.00000000000000001660
esaala	prayer			✓			500	0.0925	10	407	0.013	776.47	791.58	2.82	0.00000000000000001840
kelezia	chapel	✓	keleziya	✓			403	0.0746	4	225	0.007	776.13	791.24	3.36	0.00000000000000001842
yafeesi	office	✓	yaafeesi	✓			302	0.0559	3	87	0.003	752.21	767.32	4.32	0.00000000000000002029
zabbuli	psalms			✓			253	0.0468	2	35	0.001	748.87	763.98	5.38	0.00000000000000002057
amaingira	the process of entering	✓	amaingila		✓	**okwingila** 'to enter' is in; but not the deverbative	399	0.0738	2	238	0.008	742.55	757.65	3.27	0.00000000000000002111
okweyanza	to say thank you			✓			418	0.0773	2	278	0.009	733.34	748.45	3.11	0.00000000000000002194
okulokolwa	to become saved			✓			230	0.0425	2	23	0.001	716.16	731.27	5.84	0.00000000000000002360
eee	(interjection)				✓	(used in 1 eg of the *Eiwanika*)	452	0.0836	26	365	0.012	704.25	719.36	2.83	0.00000000000000002485
obutikitiki	seconds			✓			192	0.0355	3	9	0.000	647.25	662.36	6.94	0.00000000000000003225
rose	proper name						246	0.0455	2	59	0.002	643.33	658.43	4.58	0.00000000000000003286
ekisa	mercy			✓			481	0.0890	10	485	0.016	637.65	652.76	2.51	0.00000000000000003377
fuuti	foot			✓			207	0.0383	4	25	0.001	624.55	639.66	5.57	0.00000000000000003601
mutalya	proper name						197	0.0364	2	20	0.001	609.90	625.01	5.82	0.00000000000000003875
mulala	one (cl. 1/3)			✓			973	0.1800	33	1,961	0.063	600.66	615.76	1.51	0.00000000000000004063

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Keyword	Translation	Spelling issue?	Other spelling	*Eiwanika*?	Neologism?	Notes	Freq.	Freq. %	Texts	RC freq.	RC %	BIC	Log-likelihood	Log-ratio	Probability
amaghango	laws	✓	amawango	✓			186	0.0344	2	15	0.000	593.06	608.17	6.15	0.00000000000000004226
ganha	allow			✓			347	0.0642	4	241	0.008	591.00	606.11	3.05	0.00000000000000004272
muli	you are			✓		< SM + -li	814	0.1506	29	1,467	0.047	590.65	605.76	1.67	0.00000000000000004280
tughange	so that we allow	✓	tuwange	✓			275	0.0509	4	135	0.004	558.44	573.55	3.55	0.00000000000000005093
siraji	proper name						186	0.0344	7	25	0.001	549.42	564.52	5.42	0.00000000000000005358
meeni	and then	✓	me ni		✓	< **me** 'and then' + **ni** (focus); **me** not in	238	0.0440	6	86	0.003	546.09	561.20	3.99	0.00000000000000005460
byaruhanga	proper name						159	0.0294	2	6	0.000	542.26	557.37	7.25	0.00000000000000005581
dhaakaba	proper name						177	0.0327	2	19	0.001	541.92	557.03	5.74	0.00000000000000005591
kawala	proper name						161	0.0298	3	8	0.000	537.66	552.77	6.85	0.00000000000000005730
gavumenti	government			✓			719	0.1330	23	1,332	0.043	499.97	515.08	1.63	0.00000000000000007185
nh	and	✓	n'		✓	**ni** 'and'/(focus) is in	346	0.0640	2	318	0.010	488.51	503.61	2.64	0.00000000000000007724
m	proper name						661	0.1223	12	1,182	0.038	482.02	497.13	1.68	0.00000000000000008053
omwembi	singer			✓			193	0.0357	18	69	0.002	441.72	456.82	4.01	0.00000000000000010583
diguli	degree			✓			228	0.0422	9	123	0.004	440.08	455.18	3.41	0.00000000000000010707
mw	connective 'of'				✓	**-a** 'of' not in	583	0.1079	33	1,028	0.033	431.67	446.78	1.70	0.00000000000000011374
omukulembeze	leader			✓			292	0.0540	13	278	0.009	398.99	414.10	2.59	0.00000000000000014567
gabula	proper name					title of the current Busoga King	255	0.0472	5	210	0.007	385.50	400.61	2.80	0.00000000000000016236
pulezidenti	President			✓			193	0.0357	5	100	0.003	377.62	392.73	3.47	0.00000000000000017329
abatuukirivu	holy people	✓	abatuukilivu	✓			241	0.0446	4	189	0.006	375.78	390.89	2.87	0.00000000000000017600
ekigwaine	what is deserving			✓			145	0.0268	3	35	0.001	372.33	387.44	4.57	0.00000000000000018119
mbeeni	preferably			✓			192	0.0355	5	106	0.003	363.94	379.05	3.38	0.00000000000000019474
ziraba	proper name						117	0.0216	2	12	0.000	355.55	370.66	5.81	0.00000000000000020966
tughe	let us give; give us	✓	tuwe	✓		< SM + -wa + subjunctive	275	0.0509	3	289	0.009	345.95	361.06	2.45	0.00000000000000022867
eisakaramentu	religious teaching	✓	eisaakalamentu		✓	**saakalamentu** 'sacrament' is in; now with prefix	101	0.0187	2	3	0.000	344.24	359.35	7.60	0.00000000000000023230
nakaziba	proper name						111	0.0205	3	10	0.000	342.83	357.94	5.99	0.00000000000000023533

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Keyword	Translation	Spelling issue?	Other spelling	*Eiwanika*?	Neologism?	Notes	Freq.	Freq. %	Texts	RC freq.	RC %	BIC	Log-likelihood	Log-ratio	Probability
bagiira	they go / they follow	✓	bagiila	✓			101	0.0187	4	5	0.000	331.77	346.88	6.86	0.00000000000000026120
omwepisikoopi	bishop				✓	< Latin **episcopus** 'bishop'	107	0.0198	3	10	0.000	328.27	343.38	5.94	0.00000000000000027017
yoanne	proper name						234	0.0433	2	225	0.007	314.28	329.39	2.58	0.00000000000000031040
omusengo	offerings			✓			272	0.0503	4	316	0.010	312.84	327.95	2.31	0.00000000000000031498
meemale	before; after			✓			280	0.0518	22	337	0.011	311.89	327.00	2.25	0.00000000000000031803
abe	so that (s)he is			✓		< SM + **-ba** + subjunctive	487	0.0901	15	956	0.031	305.68	320.79	1.55	0.00000000000000033916
byona	also (pl.)			✓			496	0.0918	21	995	0.032	301.05	316.15	1.52	0.00000000000000035615
baghange	so that they can	✓	okuwange	✓			145	0.0268	4	67	0.002	295.43	310.54	3.64	0.00000000000000037830
mosese	proper name						94	0.0174	2	7	0.000	295.11	310.22	6.27	0.00000000000000037958
ekitaloodheka	that which is difficult to relay				✓	**okuloodha** 'to relay' is in; now in cl. 7 + neg. '**ta**' + stat. ext.	124	0.0229	3	40	0.001	288.89	304.00	4.15	0.00000000000000040645
kulw	because of	✓	ku lw'		✓	**-a** 'of' not in	118	0.0218	6	33	0.001	287.38	302.49	4.36	0.00000000000000041332
boss	English						107	0.0198	2	21	0.001	285.84	300.95	4.87	0.00000000000000042052
aghooni	there then	✓	awo ni	✓		< **awo** 'there' + **ni** (focus)	90	0.0166	4	6	0.000	285.51	300.62	6.43	0.00000000000000042211
lugololiro	in a straight manner				✓	**okugolola** 'to make straight' is in; deverbative is not	83	0.0154	2	3	0.000	276.68	291.79	7.31	0.00000000000000046697
kimoimo	sole; alone			✓			118	0.0218	2	44	0.001	259.99	275.10	3.95	0.00000000000000057081
olugololiro	in a straight manner				✓	**okugolola** 'to make straight' is in; deverbative is not	90	0.0166	2	12	0.000	258.43	273.54	5.43	0.00000000000000058205
aka	which belongs to				✓	**-a** 'of' not in	196	0.0363	16	196	0.006	252.61	267.72	2.52	0.00000000000000062663
omutukuvu	the one who is holy			✓			314	0.0581	5	523	0.017	243.87	258.98	1.79	0.00000000000000070249
ebyamagero	miracles	✓	ebyamagelo	✓			112	0.0207	3	46	0.001	236.62	251.73	3.81	0.00000000000000077509
masaaba	place name			✓			75	0.0139	2	5	0.000	235.40	250.51	6.43	0.00000000000000078819
kituufu	it is true				✓	only cl. 14 **obutuufu** 'truthfulness' is in; here used as an adj.	309	0.0572	20	523	0.017	234.64	249.75	1.76	0.00000000000000079660
siima	appreciate			✓			152	0.0281	2	124	0.004	225.18	240.29	2.82	0.00000000000000091125
cd	English			✓			69	0.0128	14	3	0.000	224.32	239.43	7.05	0.00000000000000092276
abagololeraku	those who raise [the hand]			✓			65	0.0120	2	1	0.000	222.99	238.10	8.54	0.00000000000000094098
omwegha	virgin			✓			180	0.0333	2	195	0.006	215.37	230.48	2.41	0.00000000000000105466

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Keyword	Translation	Spelling issue?	Other spelling	*Eiwanika*?	Neologism?	Notes	Freq.	Freq. %	Texts	RC freq.	RC %	BIC	Log-likelihood	Log-ratio	Probability
okugabana	to divide up			✓			151	0.0279	5	131	0.004	213.87	228.98	2.73	0.00000000000000107927
olukiiko	meeting			✓			350	0.0647	9	696	0.022	210.99	226.10	1.53	0.00000000000000112860
tuwe	give us			✓			124	0.0229	4	80	0.003	210.70	225.81	3.15	0.00000000000000113359
kiidhukizo	commemoration			✓			70	0.0129	2	8	0.000	203.10	218.21	5.65	0.00000000000000127984

Main abbreviations used in the Addenda

adj. = adjective
BIC = Bayesian Information Criterion
cl. = class
freq. = frequency in new Lusoga material (0.5m)
loc. = locative
RC freq. = frequency in reference corpus (3.0m)
SM = subject marker
stat. ext. = stative extension
texts = number of new texts keyword occurs in

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.