Corpus-driven Bantu Lexicography Part 3: Mapping Meaning onto Use in Lusoga

Gilles-Maurice de Schryver, *BantUGent, Department of Languages and Cultures, Ghent University, Ghent, Belgium; and Department of African Languages, University of Pretoria, Pretoria, South Africa (gillesmaurice.deschryver@UGent.be)*
and
Minah Nabirye, *BantUGent, Department of Languages and Cultures, Ghent University, Ghent, Belgium; and Department of Teacher Education and Development Studies, Kyambogo University, Kampala, Uganda (minah.nabirye@UGent.be)*

Abstract: This article is the third instalment in a trilogy of studies that deal with corpus-driven Bantu lexicography as applied to Lusoga. Having dealt with corpus-building in Part 1, and macro-structural aspects in Part 2, we now focus on the microstructure of a dictionary and in particular on the concept of Mapping Meaning onto Use. The starting point is Patrick Hanks's book chapter by the same title, which we transpose to a study of the high-frequent motion verb *-v-* in Lusoga. Our detailed analysis is as much practical as it is methodological.

Keywords: BANTU, LUSOGA, CORPUS LEXICOGRAPHY, DISTRIBUTIONAL CORPUS ANALYSIS, MAPPING MEANING ONTO USE, MEANING POTENTIALS, MOTION VERBS

Obufunze: Omutengeso gw'eitu ogukozesebwa mu namawanika w'ennimi dha Bantu. Ekitundu 3: Okukwanagania amakulu n'enkozesa mu Lusoga. Olupapula luno n'olwokusatu mu nteeko y'okulaga omusomo gw'omutengeso gw'eitu ogukozesebwa mu namawanika w'ennimi dha Bantu ogulaga omulimu ogw'akolebwa ku Lusoga. Oluvainhuma lw'okwandhula engeli eitu ly'Olusoga mu Kitundu 1 n'omuteeko gw'omutindiigo ogusinziilwaku okuzimba olukala lwa namungi w'ebigambowazo mu Kitundu 2, buti eisila liize ku kulaga ngeli amakulu g'ebigambo ye gakwanaganizibwa n'enkozesa. Omusingi gw'eisomo elilagibwa mu kitundu kino gw'ateebwawo Patrick Hanks. Ensonga enkulu dhe yataaku eisila dhilondoolebwa okusinziila ku kigelo kya namungi w'ennhingizo entabaazi *(o)ku.v.a.* Olupapula luno lugelaagelania engeli ennhingizo eno bwe yaingizibwa mu *Eiwanika ly'Olusoga* elitaasinziililwa ku itu lya bigambo n'engeli gye yandibaile esengekebwa singa eitu n'ebigelo by'emiwendo egilagibwamu byali bikozeseibwa. Eby'asoboka n'ebitaasoboka bigelaagelanhizibwa n'ebigendelelwa by'omusingi gw'eisomo ly'eitu lya namawanika mu mpandiika y'amawanika.

Ebigambo ebikulu: BANTU, LUSOGA, EITU LYA NAMAWAIKA, ENNEKEENEENIA Y'EBIGELO BY'EMIWENDO EBILAGIBWA MU ITU LY'OLULIMI, OKUKWANAGANIA AMA-KULU N'ENKOZESA, AMAKULU AGASOBOKA, KINANTABILA OMUTABAAZI

1.   Goal of the present study

In this article we wish to investigate how meaning potentials may be drawn from usages as found in a Bantu-language corpus, through an approach known as 'mapping meaning onto use' (Hanks 2002), as applied in the ongoing compilation of a new Lusoga dictionary. With this topic we are squarely dealing with a dictionary's microstructure, although the method may of course be used (and *is* used) in the field of Bantu corpus linguistics more generally, as may be seen from the recent PhDs of Nabirye (2016) for Lusoga, Kawalya (2017) for Luganda, and Mberamihigo (2014), Nshemezimana (2016) and Misago (2018) for Kirundi.

The major reference for any corpus-based microstructural issues in Bantu lexicography is de Schryver and Prinsloo (2000). In the academic literature, the attention paid to the microstructural level is far more extensive than that paid to the macrostructural level, even in articles that aim to give a perspective on both (Prinsloo and de Schryver 2001, de Schryver 2008) or in articles that take the 'lemmatisation of ...'-formula as a point of departure (de Schryver et al. 2004: 37), which is at heart macrostructural in nature but typically develops into a discussion of microstructural aspects. This may briefly be illustrated with dictionary research undertaken for Northern Sotho.

The 'lemmatisation of ...'-formula may be found in the numerous corpus-based lexicographic studies for the various word classes and other word sets of Northern Sotho, including: reflexives (Prinsloo 1992), verbs (Prinsloo 1994, Prinsloo and Gouws 1996, de Schryver and Prinsloo 2001), adjectives (Gouws and Prinsloo 1997), nouns (Prinsloo and de Schryver 1999, Bosch and Prinsloo 2002), days (de Schryver and Lepota 2001), loan words (Nong et al. 2002), copulatives (Prinsloo 2002), terms (Prinsloo and de Schryver 2002, Taljard and de Schryver 2002), adverbs (Prinsloo 2003), demonstrative copulatives (de Schryver et al. 2004), concords and pronouns (Prinsloo and Gouws 2006), and kinship terms (Prinsloo 2012, Prinsloo and Bosch 2012, Prinsloo 2014b). The opposite also occurs, namely when a primarily microstructural aspect impacts the macrostructure, again with examples for Northern Sotho: left-expanded microstructures (Gouws and Prinsloo 2005), reversibility (de Schryver 2006), communicative equivalence (Prinsloo 2006), and paradigms (Prinsloo 2014a). It has furthermore been noted that the distinction between the macrostructural and microstructural levels tends to disappear in a digital dictionary environment, as has also been illustrated abundantly for Northern Sotho (Prinsloo 2005, Prinsloo et al. 2012, Prinsloo et al. 2014, Prinsloo et al. 2017). Lastly, dictionary reviews, of for instance the corpus-based *Oxford Bilingual School Dictionary: Northern Sotho and English* (de Schryver 2007), likewise tend to focus on microstructural aspects (Prinsloo 2009, Chabata and Nkomo 2010, Faaß 2010, Klein 2010a, b, Madiba and Nkomo 2010, Kosch 2013).

While the use of a corpus to create the microstructure of a Bantu-language dictionary is thus arguably not a novel undertaking in the field, we do add to

the existing studies: (i) a theoretical framework for the current practice,[1] and (ii) a detailed analysis of how one actually goes from concordance lines to dictionary lines. In the process we will also explore two further issues, namely: (i) the differences between the use of a corpus and a manual effort, and (ii) the potential enhancement of illustrative material through the exploitation of corpus metadata.

2.   On methods and theoretical models

2.1   Corpus linguistics

The description of any language — whether in dictionaries, grammars or other reference works — should be based on real usage of that language. While one could claim that this ought to be the obvious approach, even a cursory look at much of the output by linguists shows otherwise. As adherents of the work of Patrick Hanks, we find the following quote most appropriate:

> [...] the literature of twentieth-century linguistics is strewn with examples of self-fulfilling theoretical prophecies, in which bizarre examples are first invented, then judged to be acceptable (according to the researcher's intuitions), and then presented as evidence for conclusions about some aspect of the nature of language or linguistic rules. (Hanks 2013: 307)

In order to be able to describe 'real' language,[2] large quantities of actual occurrences of that language are first collected, and then brought together in what is known as 'an electronic corpus'. Dedicated corpus-query software, such as WordSmith Tools (Scott 1996–2018), is used to search and help quantify the hard evidence found in a corpus. At that point, and only at that point, does the researcher *explain* that evidence:

> There is a huge difference between consulting one's intuitions to *explain* data and consulting one's intuitions to *invent* data. Every scientist engages in introspection to explain data. No reputable scientist (outside linguistics) invents data in order to explain it. It used to be thought that linguistics is special — that an exception could be made in the case of linguistics — but comparing the examples invented by linguists with the actual usage found in corpora shows that this is not justifiable. (Hanks 2013: 20)

To an increasing number of researchers in the language sciences the power of natural language data is compelling indeed, and for major languages this has given rise to the vibrant field of corpus linguistics, for which Sinclair (1966) may be considered the pioneering study.[3] Now half a century on, the field of corpus linguistics is booming; the *International Journal of Corpus Linguistics*, for instance, celebrated its 20th anniversary in 2015.

Crucial for corpus linguistics is to have access to a fair amount of textual data — at least a million running words, although for major languages corpora

of several billion words are not uncommon (Kilgarriff 2003–18). For languages of limited diffusion — be those minor, minority, endangered or simply neglected languages — the lack of sufficient textual data is typically the bottle-neck. Billion-word corpora are obtained by crawling the web (de Schryver 2002), a type of corpus-building effort for which most aspects are automated. Transcribing naturally-occurring speech, the default for documentary linguists, is known to be both time-consuming and costly. However, for more and more formerly under-resourced languages, written material is becoming available online (Scannell 2003–18), and for those languages the prospect of applying techniques from the field of corpus linguistics comes into view.

2.2   Bantu corpus linguistics (BCL)

The prospect of applying techniques from the field of corpus linguistics has now become a reality for a good number of Bantu languages. For Lusoga in particular, corpus-building efforts have been described in Part 1 of the present series of three articles. There it was shown that, in addition to an oral compo-nent of over half a million words in the 1.7m Lusoga corpus, about a quarter of a million words were found on the Internet, the rest of the corpus being mainly the result of the digitalisation of printed materials.

The field of Bantu corpus linguistics is about two decades old, and is reckoned to have begun with de Schryver's (1999) corpus take on the phonetics of Cilubà. Subsequently, and together with colleagues from South Africa, de Schryver effectively established BCL as a feasible research methodology. While de Schryver was at the University of Pretoria, corpus-based linguistics was undertaken for Zulu (de Schryver and Gauton 2002, Gauton et al. 2004) and for Northern Sotho (Taljard and de Schryver 2002, de Schryver and Taljard 2006). Related work was also done at the universities of Helsinki and Dar es Salaam on Swahili (Sewangi 2000, 2001, Toscano and Sewangi 2005). This early work tended to be corpus-based (i.e. studies for which a corpus is used as one source of evidence in addition to others), in contrast to more recent studies which tend to be corpus-driven (i.e. studies in which a corpus itself is considered to be the sole source of hypotheses about language) — a distinction we owe to Tognini-Bonelli (2001).

The team at the University of Pretoria has since furthered the field of BCL, as may be seen in studies on Northern Sotho (Taljard 2006, de Schryver and Tal-jard 2007, Taljard 2012, Taljard and de Schryver 2016). Meanwhile at BantUGent (i.e., the UGent Centre for Bantu Studies), an increasing number of research articles includes aspects of BCL, as seen in studies on Lusoga (de Schryver and Nabirye 2010, Nabirye and de Schryver 2011, Nabirye 2016), on Cilubà (De Kind and Bostoen 2012, Dom et al. 2015), on Kirundi (Bostoen et al. 2012, Mberamihigo 2014, Lafkioui et al. 2016, Mberamihigo et al. 2016, Nshemezimana 2016, Nshemezimana and Bostoen 2016, Devos et al. 2017, Misago 2018), on Swahili (Devos and de Schryver 2013, 2016), on Kikongo (De Kind et al. 2013, Bostoen

and de Schryver 2015, De Kind et al. 2015), and on Luganda (Kawalya et al. 2014, Kawalya 2017, Kawalya et al. 2018). Not all of these studies are truly corpus-based, let alone corpus-driven, as some of them are closer to being 'corpus-illustrated' (Tummers et al. 2005) or even tend to use their corpora as fish ponds:

> Some famous and influential linguists have simply denied the relevance of corpus evidence to linguistic theory. Others have in recent years treated corpora as 'fish ponds' in which to angle for fish that will fit independently conceived hypotheses and theories. Fish that don't fit the theory are thrown back into the pond. [Note: I owe this metaphor to John Sinclair, in conversation some years ago.] (Hanks 2013: 7, 431)

On the relationship between corpus-driven and fish-pond linguistics, Hanks furthermore points out:

> **Corpus-driven** research [...] attempts to approach corpus evidence with an open mind and to formulate hypotheses and indeed, if necessary, a whole theoretical position on the basis of the evidence found. If work is merely 'corpus-based', [Tognini-Bonelli] argues, it risks missing important insights. A truly empirical linguist (or lexicographer) is 'driven' by the data in the corpus. [... The fish pond] analogy is no doubt unfair, for even Tognini-Bonelli, Sinclair, Stubbs, Hanks, and other empirical linguists cannot avoid making some theoretical assumptions as a starting point and using examples selectively, not merely randomly. However, a corpus-driven linguist holds her or his theoretical assumptions lightly and is ready to reconsider them in the light of accumulated evidence. (Hanks 2012: 417)

Therefore, whenever possible, any future studies for Bantu languages should aim to be *driven* by corpus data. This, too, is valid for the field of lexicography, in our case for the compilation of Lusoga dictionaries.

2.3   Distributional corpus analysis (DCA)

For each aspect for which a corpus is used, a corpus analyst first takes stock of the evidence through an approach that has been termed 'distributional corpus analysis'. Geeraerts (2009: 422-423) proposes to view DCA of the Sinclair-type as a neostructuralist approach to lexical semantics, with, as its main characteristic, the 'radical usage-based rather than system-based approach: it considers the analysis of actual linguistic behaviour to be the ultimate methodological foundation of linguistics' (Geeraerts 2010: 168). Hanks, however, takes issue with Geeraerts's view of DCA as primarily a method, not a model, and comments:

> This is odd, because examination of the work of corpus analysts such as Sinclair, Hoey, Wray, Stubbs, Moon, Partington, Semino, McEnery, Hanks, and others would show that corpus analysis lends support to a model of linguistic behav-

> iour founded on prototypical usage — and Geeraerts himself is a proponent of the theory of conceptual prototypes. (Hanks 2015: 102-103)

Entering the fray on whether or not corpus linguistics is more than a methodology goes beyond the scope of the present study. It is certain, however, that in the field of Bantu lexicography, we do use DCA as a method to arrive at various distributions (of homonyms, of meaning potentials, etc.). We nonetheless also like to believe that corpus linguistics is a/our theoretical model.

2.4   Mapping meaning onto use

The various lexicographic uses of a corpus on the macrostructural level have been described, and were illustrated for Lusoga, in Part 2 of the present series of three articles. When querying a corpus in order to compile a dictionary's microstructure, there are at least five uses of that corpus: (i) to map meaning potentials, (ii) to verify and support mother-tongue intuitions, (iii) to study various distributions, (iv) as a source of examples, and (v) to provide overall counts. Working briefly through this list, from last to first, and with a focus on our Lusoga case study, we can note the following. As far as corpus counts are concerned, these are a natural by-product of the steps described in Part 2. There, it was shown that the output of the lemmatisation effort consists of 'skeleton dictionary articles', each with a lemma, part of speech, frequency, rank, frequency band and (optionally) a short meaning. The relative frequency of each candidate lemma sign is, in other words, known at the start of the compilation of each dictionary article.

Each meaning potential that will eventually be singled out is ideally also illustrated with one or more of the corpus lines that were studied to arrive at that meaning. It is a good idea to include information on the source (cf. the *Filename* in Part 1) in one way or another, with the aim to either show it overtly in 'the' or in 'one of several' final lexicographic products, or to only keep it on file for the dictionary-makers while hiding it from the target users, so that the evidence may always be traced.

As one works through the corpus lines, one is bound to begin sorting and grading the evidence, whereby one automatically ends up drawing up distributions, which may again either be used implicitly or explicitly in the actual dictionary/-ies.

Regarding intuition, it has already been pointed out that the corpus analyst needs her or his own intuition to explain data, but in order to wade through the mass of data beyond the word level, intuition is also an excellent trait to start *exploring* the corpus with. It is good to make ample use of it, but subsequently one should always stick to the principles of corpus-driven analysis in explaining the evidence. What exists is mentioned, what doesn't appear in the corpus (when expected on intuition) may or may not be pointed out. Of

course the latter does not mean that something definitely cannot occur and/or would be ungrammatical, as 'no amount of corpus evidence will provide negative evidence — evidence for what *cannot* occur' (Hanks 2013: 415). This is not a problem, as 'being able to make predictions about probable usage is much more useful than speculating about the boundaries of possibility' (Hanks 2013: 415).[4]

As regards the meaning, it may come as a surprise to non-lexicographers but it is well-known to lexicographers: no single mother-tongue speaker knows 'all the words' of her or his language (a feature lexicographers make you believe they possess; after all, aren't they supposed to say something about every word of a language?). As a matter of fact, corpus data continuously challenges what one assumes one knows about words and their meanings. Meanings, in short, can only sensibly be derived from their uses as seen in a corpus, through a principle known as Mapping Meaning onto Use (Hanks 2002), which uses the technique of Corpus Pattern Analysis (Hanks 2004), itself based on the Theory of Norms and Exploitations (Hanks 2013). Reference is made to these seminal works for the full theoretical framework. The problem has been stated by Hanks as follows:

> Existing dictionaries may be guilty of sins of omission (e.g. in accounting for pragmatics and function words), but they are equally guilty of sins of commission. They can make things seem even more complicated than they really are. In part, this is because the structure of a traditional dictionary entry is dictated by meanings not by use. Word meaning (if such a thing exists at all) is extremely vague and unstable. A word can have about as many senses as a lexicographer cares to perceive. (Hanks 2002: 159)

To which Hanks proposes the following solution:

> [...] the lexicographer must first group the corpus evidence for each word according to the contexts in which it occurs, and then decide to what extent it is possible to group different contexts together (on the grounds that they express what is essentially the same meaning), and to what extent it is necessary to make distinctions. ¶ With the advent of large corpora, it is possible to be much more precise about the typical contexts in which a word is used, and to associate different meanings with different contexts. The crucial point here is to choose, as an organizing principle for the dictionary entry, *context* (which is objectively observable and measurable) rather than *meaning* (which is opaque and depends on the perceptions of the definer). Lexicographers should think first in terms of syntax and context (or, more strictly, syntagmatics), rather than directly in terms of semantics. They can thus approach meaning indirectly, through syntagmatic analysis, according to a motivated grouping of the evidence. (Hanks 2002: 159-160)

In short, then, and with reference to our new dictionary project for Lusoga, in addition to the brief meanings as may already be logged following lemmatisation in the dictionary writing system (i.e., the TLex file (Joffe and de

Schryver 2002–18)), the main use of a corpus on the microstructural level is to say more about word meanings in context.

3.   A case study for Lusoga

3.1   Choosing the Lusoga case study

We now wish to illustrate the mapping of meaning onto use for Lusoga lexicography. Compared to working *on* English and writing about the process *in* English, which is already quite hard enough, we have the additional problem that we need to translate everything *out of* Lusoga and *into* English for the reader to be able to follow. Hanks's (2002) article on the topic, which also bears the title 'Mapping Meaning onto Use', has been summarised as follows:

> Hanks presents his own corpus analyses of *lean* and *tank* for lexicographical purposes. Rare are such detailed accounts in which the reader is led by the hand and allowed to see how the master cuts his way through the corpus vines. The latter, including their analyses, are displayed in full as addenda, hereby allowing the reader to appreciate the hesitations — about which Hanks is quite open — even more. Once the path has been cut, once Hanks unspun the hanks, the reader is offered the view that syntagmatics in tandem with 'perceived meaning' ought to be the organising principle of dictionary entries for verbs and adjectives. The organisation for nouns is similar, but slightly more complicated.
> (de Schryver 2005: 423)

In other words, just two words are used to illustrate the process, one verb (*lean*) and one noun (*tank*). For reasons of space, and given that we also need to translate our material, we will limit our current analysis for Lusoga to just one verb. For an idea of the issues involved in undertaking a study of the Lusoga noun using a corpus, see de Schryver and Nabirye (2010), which contains a section on the semantic import of the noun in Lusoga.

The Lusoga verb chosen for the present case study is the motion verb *-v-*. The root of this verb consists of just one letter, the letter 'v', which immediately indicates the additional difficulty of merely *finding* this verb in a raw corpus, thus one without any morphological analysis, which the 1.7m Lusoga corpus was before lemmatisation. We, however, took up the challenge.

3.2   The verb *-v-* in the monolingual Lusoga dictionary

To begin the discussion in a practical way, we will be employing a shortcut, by translating the relevant information gleaned from the *Eiwanika ly'Olusoga* (Nabirye 2009b), which is a monolingual dictionary of Lusoga, compiled *without* access to a corpus. This dictionary has also been digitised (Nabirye and de Schryver 2013), and is available on disc as well as freely online from http://menhapublishers.com/dictionary/. In that dictionary, the verb *-v-* is to

be found on page 379, as two homonymous forms, and as two lemma signs with the locative enclitics *-ku* and *-mu* respectively. This page is shown in Addendum 1, while the slightly edited and reformatted online data is shown in Table 1, on the left.

Table 1:   The dictionary articles for the verbs *-v-*[1], *-v-*[2], *-vaaku* and *-vaamu* in the *Eiwanika ly'Olusoga* (Nabirye 2009b), together with translations

e-Eiwanika	Translation
(o)ku.v.a[1] [(o)kúvá] *kt.[L]* *[-viile]* [nviile] bl: [Lg: okuva]	[lemma sign, part of speech, morphological information, cognate in Luganda]
1. Okusimbuka mu kifo ekilala waayolekela ekindi. gez: *Nva Mayuge.*	**1.** to depart. e.g. *I come from Mayuge.*
2. Okusibuka. gez: *Nva Iganga.*	**2.** to come from. e.g. *I hail from Iganga.*
3. Okulekela ekintu ky'obaile okola. gez: *Ebyo nabiviileku.*	**3.** to abandon. e.g. *I gave up on those things.*
4. Okuseguka mu ngila oba mu kifo. gez: *Leka nkuviile ofune eidembe.*	**4.** to make way. e.g. *Let me pave the way for you so that you get peace.*
ssk:	[combination(s) with the lemma]
● **Okuva ku luguudo***: Okwonooneka / Okuva ku mulembe*	● **to go off the road:** *to be completely damaged and unusable / to be out of fashion*
(g)gl:	[proverbs]
♦ Awava akwita n'awava akukobela	♦ The person who warns you comes from the same place as the person who will kill you
♦ Awava ennume waila nnume	♦ Where a male leaves another male will take over that place
♦ Awava mwino tiwaila mwino: Awava eliino waila ilibu	♦ The gap that your friend leaves is not filled by another friend: A gap takes the place of a tooth that has left
♦ Awava mwino tiwaila mwino: Awava eliiso waila itulu	♦ The gap that your friend leaves is not filled by another friend: Blindness takes the place of the eye that has left
♦ Akaviile mu igi tikatya ikoli	♦ The one that has just come from an egg does not fear an eagle
♦ Edhiva okulala n'embilo	♦ The steps you take one after the other develop into running
♦ Empambo eva ku kiwalo	♦ A wise lesson is learned from the cradle
♦ Ennhonhi eva ewala temala mutonto	♦ The bird that comes from far away does not finish up the edible fruit
♦ Ensanafu eva ku mugendelo telwa kufuuka kabasa	♦ The safari ant that leaves the trail does not take long to turn into a traitor
♦ Atava ku mulungi afa t'awoza	♦ The one who does not let a beautiful one alone dies while still giving explanations
♦ Ka nduviile ku ntobo oti n'omuyala atuuse we bafumba	♦ Let me start from the very beginning like the hungry person who has arrived at the place where food is being cooked
♦ Olusubi olulala we luva ku ndhu tetoonha	♦ When one blade of grass falls off the house the house does not leak
♦ Omukazi omulungi nnimilo ya ngila buli avaayo agyegwaniza	

♦ Omukwano guva mu ngabo ♦ Omukwano guva mu ngila gwatuuka eka ♦ Omusaadha kikele kiva kyonka mu bwina ♦ Va we ndi takulwania ♦ Va ku ntebe ya lata awulilila ku ise ♦ W'ova tosoile w'otela okwila ***bbgz: Okuviila, Okuviisa.***	♦ A beautiful woman is a garden along the road: whoever comes by wants it for himself ♦ Friendship comes from the shield (sharing) ♦ Friendship comes from the road and it is brought home ♦ A man is a frog, which comes out of the hole by itself ♦ (The one who says that) 'Go away from where I am' should not make you fight ♦ (The one who says that) 'Go away from father's chair' has heard it from his father ♦ The place that you leave without picking a quarrel is the one where you always return [lemma plus verbal extensions] *okuviila* [+ APPL ext.], *okuviisa* [+ CAUS ext.]
(o)ku.v.a[2] [(o)kúvá] *kt.[L] [-viile]* [nviile] bl: [Lg: okuva] Okutandiikila mu kifo ekilala okutuuka ku kifo ekindi. gez: *Ennhandha ya Nalubaale eva Idhindha.*	[lemma sign, part of speech, morphological information, cognate in Luganda] to start at a given point and move in the direction of another. e.g. *Lake Victoria starts in Jinja.*
(o)ku.v.a.a.ku [(o)kúvááku] *kt.[T] [-viileku]* [nviileku] bl: [Lg: okuvaako] **1.** Okuzima. gez: *Amasaanhalaze gaviileku.* **2.** Okuleka. gez: *Omwenge nguviileku.* / *Oyo namuvaaku naafuna owundi.* ***bbgz: Okuviilwaku, Okuviisaaku.***	[lemma sign, part of speech, morphological information, cognate in Luganda] **1.** to go/turn off. e.g. *The electricity has gone off.* **2.** to let alone/put aside. e.g. *I have put alcohol drinking aside. / I left that person alone and got another.* [lemma plus verbal extensions] *okuviilwaku* [+ APPL + PASS ext.], *okuviisaaku* [+ CAUS ext.]
(o)ku.v.a.a.mu [(o)kúváámu] *kt.[T] [-viilemu]* [nviilemu] bl: [Lg: okuvaamu] **1.** Obutatuukiliza kye wasuubiza omuntu. gez: *Tubaile tusuubiila nti agya kutuyamba aye atuviilemu.* **2.** Okulyamu olukwe. gez: *Gwetwateesa naye mwene neeyatuvaamu.* **3.** Okuwa oba okumaliliza. gez: *Bw'obifumba bivaamu bulungi.*	[lemma sign, part of speech, morphological information, cognate in Luganda] **1.** to not fulfil what is expected of you. e.g. *The one who promised to help us has failed us.* **2.** to betray. e.g. *The actual person we planned with is the one who betrayed us.* **3.** to turn out well. e.g. *When you cook them they come out very well.* **4.** to make a loss. e.g. *I have come out with nothing.* **5.** to not be properly fixed. e.g. *These shoes do not fit.*

4. Okubula ky'ofuna oba ky'ogobolola mu kintu. gez: *Nze nviiliilemu awo.* **5.** Okusagala. gez: *Eno engaito evaamu.* **bbgz: *Okuviilamu*, *Okuviisamu*.**	[lemma plus verbal extensions] *okuviilamu* [+ APPL ext.], *okuviisamu* [+ CAUS ext.]

Intuition combined with the fieldwork that led to the dictionary data seen in Table 1 clearly indicate that the verb(s) -*v*-, without and with locative enclitics, is/are indeed quite polysemous.

3.3   The verb -*v*- in the Lusoga lemmatised frequency list

From the 1.7m Lusoga corpus (cf. Part 1), a lemmatised frequency list was created (cf. Part 2). Perusing it, we notice that the data for the verbal lemma -*v*- was not split into two. Deciding whether or not to create two homonyms for -*v*- was not feasible during lemmatisation, where the focus was literally on lemmatisation and part-of-speech assignment, not on any detailed studies of usage leading to meaning. When it comes to the verbal forms with locative enclitics, however, we find not just -*vaaku* (with an enclitic from cl. 17) and -*vaamu* (cl. 18) in the lemmatised frequency list, but also -*vaawo* (cl. 16) and -*vaayo* (cl. 23). From a frequency point of view, then, one can say that the latter two locativised verbs were 'overlooked' during the manual (i.e., non-corpus) effort to compile the monolingual Lusoga dictionary. Also overlooked in the *Eiwanika ly'Olusoga* is the deverbative noun -*vo* in cl. 14, which does have a respectable frequency in the lemmatised frequency list. These six lemmas are listed in Table 2, together with their lemma frequencies, lemma ranks, lemma frequency bands, as well as number of formatives.

Table 2:   The lemmas -*v*-, -*vaawo*, -*vaaku*, -*vaamu*, -*vaayo* and cl. 14 -*vo* in the lemmatised frequency list derived from the 1.7m Lusoga corpus

Lemma	Part of speech	Freq.	Rank	Freq. band	# formatives
-v-	*verb*	6 611	21	①	67
-vaawo	*locativised verb (cl. 16)*	15	3 679	-	1
-vaaku	*locativised verb (cl. 17)*	14	3 852	-	1
-vaamu	*locativised verb (cl. 18)*	242	571	②	8
-vaayo	*locativised verb (cl. 23)*	281	518	②	23
-vo	*deverbative noun, in cl. 14*	40	2 096	-	2

The formative (or underlying) data that led to the six lemmas listed in Table 2 is presented in Addendum 2. For the verb -*v*-, for instance, 67 types were

frequent enough — meaning that their frequency was at least 12 in the 1.7m Lusoga corpus (cf. Part 2, §3) — and the frequencies of these 67 all contribute to the total frequency of the lemma -*v*-, being 6 611, which turns out to be one of the most frequent lemmas in the language, with rank 21. From Table 2 one may further conclude that given that -*vaaku* was entered in the *Eiwanika ly'Olusoga*, -*vaawo* with a similar frequency and cl. 14 -*vo* should indeed have been entered as well, and especially the top-frequent -*vaayo*, the 518th-most-frequent lemma overall in Lusoga.[5]

3.4   The verb -*v*- in the 1.7m Lusoga corpus

3.4.1   Mapping steps and sampling procedure

We are now in a position to study the Lusoga corpus evidence for -*v*-. The steps of the procedure to map meaning onto use have been enumerated as follows by Hanks, with reference to his case study of English *lean*:

> Working with a 500-line sample, we sort all the occurrences into different categories, first on broad syntactic grounds (separating adjectives from the verbs), then into more delicate semantic and syntactic frames (e.g. separating 'lean meat' from 'lean businesses') and finally making more subtle distinctions on semantic grounds (e.g. separating different meanings of 'lean on someone', according to the perceived purpose of the person doing the leaning, i.e. reliance or choice). [...] It should be emphasized that the level of detail used in categorization of corpus lines is a matter of choice and judgement: even more delicate subcategorization is possible, or different patterns may be lumped together in a single category. (Hanks 2002: 165-166, our underlining)

Without any further information, sampling the raw Lusoga corpus in search of -*v*- is obviously hard. However, once one realises that one has the underlying forms which led to each lemma at hand, the process is actually perfectly doable. According to the data presented in Addendum 2, the most frequent formatives for the lemma -*v*- are *okuva* (freq. 2 668), *ava* (freq. 389), *ova* (freq. 325), *kuva* (freq. 267), *yava* (freq. 188), *nva* (freq. 162), etc. In other words, one may simply instruct WordSmith Tools to search for any or all of such frequent types at the same time (by simply placing slashes between the various forms), with or without a randomiser (for instance, to limit the output to a sample of 100 lines), to then study the concordance lines. As an alternative, adding a verbal extension, such as an applicative, or the perfect, and searching for -*viil*- rather, is also an option.

3.4.2   The verbs -*v*-¹, -*v*-², the connective *kye*-SM-*va*, and the adverb *kuva*

After a careful study of several hundreds of concordance lines for -*v*-, we concluded that the various uses are indeed best presented in two separate,

homonymous, dictionary entries. Given that we are describing the evidence in English, there may be a tendency to let the English categories influence the Lusoga evidence. We have avoided that, just as it is good practice in bilingual lexicography not to allow the target language to 'pull' or 'distort' the source language analysis (Atkins 1996: 8).

The various verbal uses as seen in the corpus lead to the meaning potentials listed below, ordered from more to lesser frequent, and grouped around usages that have to do with movement, vs. usages that have to do with projection and direction. Adding an addendum with the many concordance lines will not be beneficial to the reader; instead, we add a glossed example for each use. (For the abbreviations in the glosses, see the explanations at the end.)

okuva¹ [move senses]

1. to leave, to depart, to go away
2. to hail (from)
3. to abandon
4. to make way, to move away
5. to result, to come out
6. to spend (time)

1.   to leave, to depart, to go away
*Kuba ye bwe **yava** e Makeerere nga amaze okufuna diguli, yafulumamu bufulume yaagya ku mawanga.*

kuba	ye	bwe	**a-a-v-a**		e	makeerere	nga
because	him	when	**SM₁-PST-leave-FV**		LOC₂₃	9.Makerere	CON
a-mal-ile		oku-fun-a	diguli		a-a-fulum-a-mu		
SM₁-finish-PERF		15-get-FV	9.degree		SM₁-PST-exit-FV-ENCL₁₈		
bu-fulum-e		a-a-gi-a		ku	ma-wanga		
14-exit-DEV		SM₁-PROG-go-FV		LOC₁₇	6-tribe		

'Because for him when **he left** Makerere after getting his degree, he just left and went abroad.'
[File ID: PFExtaud | O • Interviews • Language • 2012]

2.   to hail (from)
Oviiliile wa mu bufunze?

o-v-il-ile		wa	mu	bu-funz-e
SM₂SG-hail-APPL-PERF		INTER	LOC₁₈	14-brief-DEV

'Where **did you hail from**, in brief?'
[File ID: Ebintub | O • Songs - Traditional • Life • 2010s]

3.   to abandon
*Omuntu bwe yeetukuza n'**ava** mu bibi byonabyona, afuuka ekibya ekikozesebwa emilimo egya ghaigulu.*

o-mu-ntu   bwe a-e-etukul-a   ni   **a-v-a**   mu
AUG-1-person if   SM₁-RFL-clean-FV   CON **SM₁-abandon-FV** LOC₁₈
bi-bi   bi-ona-bi-ona   a-fuuk-a   e-ki-bya
AP₈-bad   PP₈-INC-PP₈-INC   SM₁-transform-FV   AUG-7-plate
eki-koz-is-ibw-a   e-mi-lim-o   e-gi-a   ghaigulu
SREL₇-use-CAUS-PASS-FV   AUG-4-work-DEV   AUG-CP₄-of   above
'If a person becomes holy and **s/he abandons** all forms of sinful states, s/he becomes a vessel that can be used to do jobs of a high rank.'[6]
[File ID: Endagaan | W • Biblical documents • Religion • 1998]

4.   to make way, to move away
Nva ni mu maiso tutasambaganilagho [...]
n-v-a   ni   mu   ma-iso
SM₁SG-move_away-FV   even   LOC₁₈ 6-eye
tu-ta-samb-agan-il-a-gho
SM₁PL-NEG_B-kick-REC-APPL-FV-ENCL₁₆
'**Move away** even from my presence; let us not kick each other from here [...]'
[File ID: AkatAkas | W • Literature • Fables • 1999]

5.   to result, to come out
Ebiviile mu kubuuzibwa [...]
ebi-v-ile   mu   ku-buuz-ibw-a
SREL₈-result-PERF   LOC₁₈   15-question-PASS-FV
'**What came out** of the examination [...]'
[File ID: Missa4 | W • Biblical documents • Religion • 2012]

6.   to spend (time)
Aye nga Abadiope baviile ekiseela nga nga beesabila ela nga bemba nti oba twena tuliba awo twafuna ku masaanhalaze.
aye_nga   a-ba-diope   **ba-v-ile**   e-ki-seela   nga   nga
but   AUG-2-diope **SM₂-spend-PERF** AUG-7-period ADV   ADV
ba-e-sab-il-a   ela   nga   ba-emb-a   nti   oba
SM₂-RFL-request-APPL-FV CON   ADV   SM₂-sing-FV that   MOD
tu-ena   tu-li-b-a   a-wa-o   tu-a-fun-a
SM₁PL-INC   SM₁PL-FUT₂-be-FV   AUG-PP₁₆-DEM_B   SM₁PL-PROG-get-FV
ku   ma-saanhalaze
LOC₁₇   6-electricity
'But the Badiope **have spent** a long time nagging so that we might also be there (one day) and we (finally) get a bit of electricity.'
[File ID: Mwino | O • Radio talk shows • Politics • 2010]

okuva² [*projection and direction senses*]

1.   to start (and continue onwards)
2.   to be the source (of), to emanate (from)

1.   to start (and continue onwards)

Ate **kiviila** *ilala ku Bbaibbuli wano Yesu bwe yagyanga nga alonda abayigilizwa yaabaagaananga na ki? Ni profession dhaibwe.*

ate	**ki-v-il-a**	ilala	ku	bbaibbuli	wa-no	yesu
CON	**SM$_7$-start-APPL-FV**	INTENS	LOC$_{17}$	9.Bible	PP$_{16}$-DEM$_A$	1.Jesus

bwe	a-a-gi-ang-a	nga	a-lond-a	a-ba-yigilizwa
ADV	SM$_1$-PST-go-HAB-FV	ADV	SM$_1$-pick-FV	AUG-2-disciple

a-a-ba-agaan-ang-a		na	ki	ni	profession
SM$_1$-PST-OM$_2$-find-HAB-FV		CON	INTER	CON	10.profession

dhi-a-ibwe
CP$_{10}$-of-POSS$_{2PL}$

'And **it** really **starts from** the Bible here when Jesus used to go and pick disciples, he usually found they were with what? With their own professions.'
[File ID: Luthour | O • Radio talk shows • Religion • 2010]

2.   to be the source (of), to emanate (from)

NBS ni radio ekutuusaaku amawulile agaba gaakagwawo ate **okuviila** *ilala ku bantu abatuufu beene.*

NBS	ni	radio	e-ku-tuus-a-ku		a-ma-wulile
9.NBS	COP	9.radio	SREL$_9$-OM$_{2SG}$-bring-FV-ENCL$_{17}$		AUG-6-news

aga-b-a	ga-aka-gw-a-wo		ate	**oku-v-il-a**
SREL$_6$-be-FV	SM$_6$-APERF-fall-FV-ENCL$_{16}$		CON	**15-emanate-APPL-FV**

ilala	ku	ba-ntu	aba-tuuf-u	ba-ene
INTENS	LOC$_{17}$	2-person	SREL$_2$-right-DEV	PP$_2$-RFL

'NBS is the radio that brings you fresh news and on top of that **emanating from** the real right people.'
[File ID: Mazima | O • Radio talk shows • Politics • 2010]

Combinations

Three combinations appear frequently in the concordance lines, the first derived from -*v*-¹, sense 1.

-v- + LOC + maiso = to die

Nga kitalo muna omukaile **okutuva ku maiso***!*

nga	kitalo	mu-na	o-mu-kaile	**oku-tu-v-a**	**ku**	**ma-iso**
as	sad	1-AFP	AUG-1-old_person	**15-OM$_{1PL}$-leave-FV**	**LOC$_{17}$**	**6-eye**

'As it is sad my counterpart for the old person **to die**!'
[File ID: Byaif12 | W • E-mails • Networking • 2012]

The next two frequent combinations are derived from *-v-*[2], sense 1, and have to do with measuring, either space or time.

-v- + -tuuk- = from ... up to [measuring space]
*[...] n'abali kwonoona **kuva** itale ghano ghati **okutuukila** ghano.*

ni	aba-li	ku-yonoon-a	**ku-v-a**	itale	gha-no
COP	SREL[2]-be	15-spoil-FV	**15-start-FV**	above	PP[16]-DEM[A]

ghati	**oku-tuuk-il-a**	gha-no
here	**15-reach-APPL-FV**	PP[16]-DEM[A]

'[...] they are the ones who are spoiling **from** above here like this **up to** here.'
[File ID: Okukyal | O • Celebrations • Politics • 2011]

-v- + paka = from ... up to, from ... until [measuring time]
*Nkola **kuva** saawa ina **paka** musanvu ogw'obwile.*

n-kol-a	**ku-v-a**	saawa	ina	**paka**	musanvu	o-gu-a
SM[1SG]-work-FV	**15-start-FV**	9.time	ten	**up_to**	3.one	AUG-CP[3]-of

obu-ile
14-night

'I work **from** 10:00 **until** 1:00 o'clock in the night.'
[File ID: BuwaabGr | O • Celebrations • Inspirational • 2010]

Other word classes

Addendum 2 indicates that, among the formatives of the verb *-v-*, one also finds the forms *kyava*, *kyebaava*, *kyenva*, *kyetuva* and *kyeyava*. These words actually belong to a different word class, as these are connectives which are built according to a fixed formula, combining the object relative of class 7, followed by a subject marker, and then *-v-*[1], sense 5.

kye-SM-va (connective) = that is why
*Buti **kyenva** tyayenze kufuna batoototo kuba boona baidha kuba bakyotala [...]*

buti	**kye-n-v-a**	ti-a-yend-ile	ku-fun-a
now	**OREL[7]-SM[1SG]-result-FV**	NEG[A]-PST-want-PERF	15-get-FV

ba-toototo	kuba	ba-ona	ba-idh-a	ku-b-a	ba-kyotala
AP[2]-young	because	PP[2]-INC	SM[2]-come-FV	15-be-FV	2-half-caste

'Now **that is (the reason) why I** did not want to get young ones because even they will be half-castes [...]'
[File ID: PFExtaud | O • Interviews • Language • 2012]

From *-v-*[2], sense 1, the adverb *kuva* is derived.

kuva (adverb) = since
*Nnhweileku aye **kuva** nkyo nkaali kulyaku.*

n-nhw-ile-ku		aye	**ku-v-a**	n-kyo	n-kaali
SM1SG-drink-PERF-ENCL17		but	**15-start-FV**	9-morning	SM1SG-not

ku-li-a-ku
15-eat-FV-ENCL17

'I have drunk a bit but **since** morning I have not eaten at all.'
[File ID: AgakbOmu | W • Literature - Novels • Life • 2012]

3.4.3   The locativised verb *-vaawo*

When the class 16 locative enclitic *-wo* is suffixed to the base verb *-v-*[1], a new use that was not seen for the base verb is found (1. below), together with the main use as also seen for the base verb (2. below).[7]

okuvaawo < okuva[1]

1.   to stop existing, to die (out)
2.   to leave, to depart, to go away

1.   to stop existing, to die (out)
Oyenda toyenda eliyo ebiidha okuvaawo.

o-yend-a	ti-o-yend-a	e-li-yo	ebi-idh-a
SM2SG-want-FV	NEGA-SM2SG-want-FV	SM23-be-ENCL23	SREL8-come-FV

oku-v-a-wo
15-die_out-FV-ENCL16

(Whether) you want or do not want, there are things which will **die out**.
[File ID: Musoke | O • Radio talk shows • Marriage • 2010]

2.   to leave, to depart, to go away
Bakaile baife abalungi mutavaawo tuli kwila.

ba-kaile	ba-a-ife	aba-lungi	**mu-ta-v-a-wo**
2-parent	CP2-of-PERS1PL	SREL2-good	**SM2PL-NEGB-go_away-FV-ENCL16**

tu-li   ku-il-a
SM1PL-be   15-come_back-FV

'Our good elders, **do not go away**, we are coming back.'
[File ID: Luthour | O • Radio talk shows • Religion • 2010]

3.4.4   The locativised verb *-vaaku*

When the class 17 locative enclitic *-ku* is suffixed to the base verb *-v-*[1], numerous new uses that were not seen for the base verb are found (all but one below), together with one main use as also seen for the base verb (2. below).

okuvaaku < okuva[1]

1. to go off, to turn off
2. to abandon
3. to trigger, to cause
4. to let aside, to give up
5. to lose
6. to stop
7. to not disturb, to leave alone
8. to finish
9. to come a (little) bit

1. to go off, to turn off
Eeh! Amasaanhalaze gaviileku [...]
eeh   a-ma-saanhalaze   ga-v-ile-ku
INTERJ   AUG-6-electricity   SM$_6$-go_off-PERF-ENCL$_{17}$
'Eeh! The electricity has gone off [...]'
[File ID: PFExtaud | O • Interviews • Language • 2012]

2. to abandon
Oba ti na kindi okutuviilaku ilala [...]
oba   ti-na-ki-ndi   oku-tu-v-il-a-ku
or   NEG$_A$-MODF-PP$_7$-EXC   15-OM$_{1PL}$-abandon-APPL-FV-ENCL$_{17}$
ilala
INTENS
'Or perhaps even to abandon us completely [...]'
[File ID: StarEC3 | O • Radio talk shows • Health • 2010]

3. to trigger, to cause
Ekiliiviilaku baana okutuluguunhizibwa nnankani, buvunaanhizibwa bwaife.
eki-li-∅-v-il-a-ku   ba-ana
SREL$_7$-PROG-15-cause-APPL-FV-ENCL$_{17}$ 2-child
oku-tuluguunh-is-ibw-a   nnankani bu-vunaanhizibwa bu-a-ife
15-abuse-CAUS-PASS-FV   UNS   14-responsibility   CP$_{14}$-of-PRON$_{1PL}$
'What is causing the children to be abused, (is) something (that is) our respon-
sibility.'
[File ID: Pl101216 | O • Radio talk shows • Health • 2010]

4. to let aside, to give up
Mwana wange, kulwaki tobivaaku?
mu-ana   a-a-nge   ku-lu-a-ki
1-child   CP$_1$-of-POSS$_{1SG}$ PP$_{17}$-CP$_{11}$-of-INTER
ti-o-bi-v-a-ku
NEG$_A$-SM$_{2SG}$-OM$_8$-give_up-FV-ENCL$_{17}$

'My child why **don't you give them up**?'
[File ID: AbabitAb | W • Literature • Fables • 1999]

5. **to lose**

Tusaasilaku abantu **abaaviilwaku** *abantu baibwe modulo oti n'eyo.*
tu-saasil-a-ku   a-ba-ntu
SM1PL-sympathise-FV-ENCL17   AUG-2-person
a-ba-a-v-il-w-a-ku   a-ba-ntu   ba-a-ibwe
AUG-SM2-PST-lose-APPL-PASS-FV-ENCL17 AUG-2-person   CP2-of-POSS2PL
modulo   oti   ni   e-e-o
9.model   like   COP   AUG-23-DEMB
'We sympathise a bit with the persons **who have lost** their people in a manner like that one.'
[File ID: StarEC2 | O • Radio talk shows • Health • 2010]

6. **to stop**

ARVs **waadhivaaku***?*
ARVs **o-a-dhi-v-a-ku**
10.ARV **SM2SG-PST-OM10-stop-FV-ENCL17**
'**Did you stop** (taking) the ARVs?'[8]
[File ID: StarEC4 | O • Radio talk shows • Health • 2010]

7. **to not disturb, to leave alone**

Nvaaku iwe akavubuka ye ggu lwaki onneesimbamu engeli eyo?
n-v-a-ku   iwe   a-ka-vubuka   ye   ggu
OM1SG-leave_alone-FV-ENCL17 PERS2SG AUG-12-youth INTERJ INTERJ
lwaki   o-n-e-simb-a-mu   e-n-geli   e-yi-o
INTER   SM2SG-OM1SG-RFL-stand-FV-ENCL18 AUG-9-way   AUG-PP9-DEMB
'**Leave me alone** you young boy; but really why do you stand against me in that way?'
[File ID: Abantub | O • Songs - Traditional • Rehabilitation • 2000s]

8. **to finish**

Bamaama bwe banaaba **baviileku** *twidha kuba n'ebigambo okuva eli babbaabba.*
ba-maama   bwe   ba-naa-b-a   **ba-v-ile-ku**
2-mother   MOD   SM2-FUT1-be-FV   **SM2-finish-PERF-ENCL17**
tu-idh-a   ku-b-a   ni   e-bi-gambo   oku-v-a
SM1PL-come-FV   15-be-FV   CON   AUG-8-word   15-emanate-FV
e-li   ba-bbaabba
SM23-be   2-father
'After the mothers **have finished** (giving their speeches), we shall be with the words emanating from the fathers.'
[File ID: BuwaabGr | O • Celebrations • Inspirational • 2010]

9.   to come a (little) bit

Nga bwe twabategeeziiza okuva eila nti munange tugya kuba n'abakungu ==okuvaaku==
==mu Judicial Service Commission.==

nga	bwe	tu-a-ba-tegeez-is-a	oku-v-a	eila	nti
like	as	SM_{1PL}-PST-OM_{2PL}-inform-CAUS-FV	15-start-FV	already	that

mu-na-nge	tu-gi-a	ku-b-a	ni	a-ba-kungu
1-AFP-$POSS_{1SG}$	SM_{1PL}-go-FV	15-be-FV	CON	AUG-2-specialist

oku-v-a-ku	mu	judicial_service_commission
15-come_a_bit-FV-$ENCL_{17}$	LOC18	9.Judicial_Service_Commission

'As we already informed you, my friends, we shall be with specialists who
==come a little bit== from the Judicial Service Commission.'
[File ID: Judicial | O • Radio talk shows • Sensitization • 2010]

Combinations

Together with the noun *omusolo* 'tax', sense 4 acquires a specific use, as shown
below.

-vaaku omusolo **= to remit tax**
Amakolelelo ==agavaaku omusolo==.

a-ma-kol-ilil-o	**aga-v-a-ku**	**o-mu-solo**
AUG-6-do-RPT-DEV	**$SREL_6$-let_aside-FV-$ENCL_{17}$**	**AUG-3-tax**

'Factories ==that remit tax==.'
[File ID: AVATVAT | O • Songs - Traditional • Sensitization • 2000s]

3.4.5   The locativised verb *-vaamu*

When the class 18 locative enclitic *-mu* is suffixed to the base verb *-v-*[1], numer-
ous new uses that were not seen for the base verb are found (3. to 5. below),
together with variations of the two main uses as also seen for the base verb
(1. and 2. below).

okuvaamu < okuva[1]

1.   to abandon though it is expected
2.   to come out, to flow out, to exit
3.   to grow well, to turn out well
4.   to yield, to generate
5.   to not gain

1.   to abandon though it is expected
Yaasangulawo bile ebibaile bili kwogelwa nti akalulu ==akaviilemu== *[...]*

a-a-sangul-a-wo	bi-le	ebi-b-a-ile	bi-li
SM_1-PST-rub-FV-$ENCL_{16}$	PP_8-DEM_C	$SREL_8$-be-FV-PERF	SM_8-PROG

ku‘ogel-w-a   nti   a-ka-lulu
15-speak-PASS-FV   CON   AUG-12-vote
a-ka-v-ile-mu
SM₁-OM₁₂-abandon_though_it _is expected-PERF-ENCL₁₈
'He cancelled all the things that have been said and as to the election <mark>he abandoned it though it was expected of him</mark>.'
[File ID: AEGY3 | O • Radio talk shows • Health • 2010]

2.   to come out, to flow out, to exit
Owundi ku baisilukale yaamufumita eifumo mu lubavu ela mangu ago <mark>mwavaamu</mark> omusaayi n'amaadhi.
o-wu-ndi-ku   ba-isilukale a-a-mu-fumit-a   e-i-fumo
AUG-PP₁-EXC-ENCL₁₇ 2-soldier   SM₁-PST-OM₁-pierce-FV AUG-5-spear
mu   lu-bavu ela   mangu   a-ga-o   **mu-a-v-a-mu**
LOC₁₈ 11-rib   CON quickly AUG-PP₆-DEM_B **SM₁₈-PST-come-FV-ENCL₁₈**
o-mu-saayi   ni   a-ma-adhi
AUG-3-blood   CON   AUG-6-water
'One of the soldiers pierced him with a spear in the rib and quickly (thereafter) blood and water <mark>flew out</mark>.'
[File ID: Missa1 | W • Biblical documents • Religion • 2012]

3.   to grow well, to turn out well
Bw'ozaala n'abaawo <mark>ekivaamu</mark> agasa bwa iwanga.
bwe   o-zaal-a   ni   a-b-a-wo
if   SM₂SG-give_birth-FV CON SM₁-be-FV-ENCL₁₆
eki-v-a-mu   a-gas-a   bwa   i-wanga
SREL₇-turn_out_well-FV-ENCL₁₈   SM₁-benefit-FV   just   5-country
'If you give birth and <mark>s/he turns out well</mark>, s/he just benefits the nation.'
[File ID: Bwozaal | O • Songs - Traditional • Inspirational • 2000s]

4.   to yield, to generate
Ekiighulo kya Kyabazinga kyatundibwa aghalala ni Kodh'eyo <mark>kyavaamu</mark> emitwalo kumpi ikumi nga empiiya dhino dha kugheeleza Busoga mu bitundu ebili n'eby'etaago.
e-ki-ighul-o   ki-a   kyabazinga ki-a-tund-ibw-a   aghalala
AUG-7-food-DEV   CP₇-of 1a.King   SM₇-PST-sell-PASS-FV together
ni   kodh'eyo   **ki-a-v-a-mu**   e-mi-twalo   kumpi
CON 9.Kodh'eyo **SM₇-PST-yield-FV-ENCL₁₈** AUG-4-ten_thousand near
ikumi   nga   e-n-piiya   dhi-no   dhi-a   ku-gheel-is-a
ten   CON AUG-10-money   PP₁₀-DEM_A   CP₁₀-of   15-serve-CAUS-FV
bu-soga mu   bi-tundu   ebi-li   ni   e-bi-etaag-o
14-soga LOC₁₈ 8-part   SREL₈-be CON   AUG-8-need-DEV
'(Tickets to attend) the 'dinner' (in honour) of the King were sold together with *Kodh'eyo*[9] and <mark>this yielded</mark> roughly one hundred thousand Shillings which serves Busoga in the parts which have needs.'
[File ID: Kodh'eyo | W • Journalism • Networking • 1997–1998]

5.   to not gain

Muzeeyi iwe oidha kuviilamu awo.

mu-zeeyi	iwe	o-idh-a	ku-v-il-a-mu
1-old	PRON_{2SG}	SM_{2SG}-will-FV	15-not_gain-APPL-FV-ENCL_{18}

a-wa-o
AUG-PP_{16}-DEM_{B}

'Mzee you, you will **not gain** anything at all.'

[File ID: PFExtaud | O • Interviews • Language • 2012]

Combinations

Together with the noun *enda* 'stomach', sense 2 acquires a specific use, as shown below.

-vaamu enda **= to miscarry**

[...] mukazi wo lw'avaamu enda. Onaagya waalima?

mu-kazi	a-o	lwe	a-v-a-mu	en-da
1-wife	PP_{1}-DEM_{B}	OREL_{11}	SM_{1}-come_out-FV-ENCL_{18}	10-stomach

o-naa-gi-a	o-a-lim-a
SM_{2SG}-FUT_{1}-go-FV	SM_{2SG}-PROG-dig-FV

'[...] (the day) when your wife **miscarries**. Will you go and dig?'

[File ID: Esaalmk1 | O • Radio talk shows • Religion • 2010]

Other word classes

One particular frequent construction has lexicalised and is used as a connective — namely the subject relative of cl. 7, with the past tense marker, and sense 2 of *-vaamu* — as shown below.

ekyavaamu **(connective) = what came out of it, what resulted from it**

Ekyavaamu, Wampala yaakoogha okutambula nga bwayagala; Wakayima, yaasalawo okugya okwekweka mu ndhu.

eki-a-v-a-mu	wa-mpala	a-a-koogh-a	oku-tambul-a
SREL_{7}-PST-come_out-FV-ENCL_{18}	16-lion	SM_{1}-PST-tire-FV	15-walk-FV

nga	bwe	a-a-gal-a	wa-kayima	a-a-sal-a-wo
ADV	ADV	SM_{1}-PROG-search-FV	16-monkey	SM_{1}-PST-decide-FV-ENCL_{16}

oku-gi-a	oku-e-kwek-a	mu	n-dhu
15-go-FV	15-RFL-hide-FV	LOC_{18}	9-house

'**What resulted from it**, (is that) Mr. Lion got tired of walking as he searched; Mr. Monkey decided to go and hide himself in the house.'

[File ID: MwidTufm | W • Literature • Fables • 1999]

3.4.6   The locativised verb *-vaayo*

When the class 23 locative enclitic *-yo* is suffixed to the base verb *-v-*[1], either a variation of sense 5 of the base verb is seen, or a new one.

okuvaayo < okuva[1]

1.   to come out
2.   to give way

1.   to come out
Abawala muviileyo? Mwanguyeeku mwanguyeeku.
a-ba-wala   **mu-v-ile-yo**
AUG-2-girl   **SM_{2PL}-come_out-PERF-ENCL_{23}**
mu-angu-y-e-ku   mu-angu-y-e-ku
SM_{2PL}-hurry_up-CAUS-SUBJ-ENCL_{17}   SM_{2PL}-hurry_up-CAUS-SUBJ-ENCL_{17}
'Girls **have you come out**? Hurry up, hurry up.'
[File ID: IntHadij | O • Celebrations • Marriage • 2008]

2.   to give way
Eee! Oooh! Vaayo baidhaakuniina.
eee   oooh   **v-a-yo**   ba-idh-a   ∅-ku-niin-a
INTERJ   INTERJ   **give_way-FV-ENCL_{23}**   SM_2-will-FV   15-OM_{2SG}-step-FV
'Eee! Oooh! **Give way**, they will step on you.'
[File ID: PFExtaud | O • Interviews • Language • 2012]

3.4.7   The cl. 14 deverbative noun *-vo*

While all previous derivations (§§3.4.3–3.4.6) were derived from *-v-*[1], one frequent derivation, the cl. 14 deverbative noun *-vo*, is derived from *-v-*[2], sense 1, as shown below.

obuvo = the beginning < okuva[2]

Inhonhola obuvo n'obwiko bwe ensonga yaali kwogelaku.
inhonhol-a   **o-bu-v-o**   ni   o-bu-ik-o   bu-a
explain-FV   **AUG-14-start-DEV**   and AUG-14-end-DEV   CP_{14}-of
e-n-songa   ye   a-li   ku-yogel-a-ku
AUG-9-issue   OREL_9   SM_1-be   15-speak-FV-ENCL_{17}
'Explain the **beginning** and the end of the issue that s/he is talking about.'
[File ID: Omugole | W • Literature - Plays • Marriage • 2007]

3.4.8   Summary of the corpus evidence for the Lusoga verb -*v*-

The corpus evidence as analysed and illustrated in §3.4.2 through §3.4.7 can now be synthesised as presented in Table 3. The three steps of Hanks's procedure may be recognised, but for a Bantu language the approach is not as linear as suggested in §3.4.1 for English. Part of Step 1, the division 'on broad syntactic grounds', is the outcome of the lemmatisation, which resulted in the distinction between verbal, locativised verbal and nominal uses (column 1 in Table 3). The other half, with connectives and an adverbial use, was only revealed during analysis (column 4 in Table 3). When it comes to Step 2, the division 'into more delicate semantic and syntactic frames' is what we termed *combinations* (column 3 in Table 3). In our case study, these may be combinations of verb + noun, verb + verb, verb + preposition, and verb + locative + noun. Those that include a preposition also turn into prepositional uses. Due to the structure of Bantu languages, some of these lemmas and combinations include codes for entire paradigms (here LOC = any locative, SM = any subject marker). Lastly, Step 3, 'making more subtle distinctions on semantic grounds', goes to the heart of the splitting vs. lumping decisions that every lexicographer must contend with (column 2 in Table 3).

Table 3:   Synthesis of the verb -*v*- in the 1.7m Lusoga corpus, with columns 1 and 4 for Step 1, column 3 for Step 2, and column 2 for Step 3 of the procedure to map meaning onto use. (Manual effort between [ ].)

Lemma signs derived from the lemmatised frequency list	Meaning potentials	Combinations, + meaning potentials following '='	Lemma signs for other word classes, + meaning potentials following '='
okuva¹ [move senses]	1. to leave, to depart, to go away [1]	-*v*- + LOC + *maiso* = to die	
	2. to hail (from) [2]		
	3. to abandon [3]		
	4. to make way, to move away [4]		
	5. to result, to come out		*kye*-SM-*va* (connective) = that is why
	6. to spend (time)		
okuva² [projection and direction senses]	1. to start (and continue onwards) [1]	-*v*- + -*tuuk*- = from ... up to [measuring space]	*kuva* (adverb) = since
		-*v*- + *paka* = from ... up to, from ... until [measuring time]	

	2. to be the source (of), to emanate (from)		
okuvaawo < okuva¹	1. to stop existing, to die (out)		
	2. to leave, to depart, to go away		
okuvaaku < okuva¹	1. to go off, to turn off [1]		
	2. to abandon		
	3. to trigger, to cause		
	4. to let aside, to give up [2]	*-vaaku omusolo* = to remit tax	
	5. to lose		
	6. to stop		
	7. to not disturb, to leave alone		
	8. to finish		
	9. to come a (little) bit		
okuvaamu < okuva¹	1. to abandon though it is expected [1]		
	2. to come out, to flow out, to exit [5]	*-vaamu enda* = to miscarry	*ekyavaamu* (connective) = what came out of it, what resulted from it
	3. to grow well, to turn out well [3]		
	4. to yield, to generate		
	5. to not gain [4]		
okuvaayo < okuva¹	1. to come out		
	2. to give way		
obuvo < okuva²	the beginning		

3.5   Comparison of the manual effort vs. the corpus evidence for the Lusoga verb *-v-*

Any comparison between a manual effort and a corpus-driven one is always unfair, as the corpus tends to 'win'. In doing so, one often forgets about the heroic efforts that went into the manual effort in the first place (Nabirye 2008, 2009a, Nabirye and de Schryver 2010, 2011, 2013). The following, therefore, is only for illustrative purposes.

While the lemmatisation had already revealed that two of the four loca-tivised verbs had accidentally been overlooked, including a very frequent one,

as well as a deverbative noun (probably because it was assumed to belong to the grammar rather than the lexicon), all of the trickier derived word classes as well as the truly frequent combinations were also absent from the manual effort. (The one combination offered in the monolingual dictionary, viz. *okuva ku luguudo*, was not found in the 1.7m Lusoga corpus.) With regard to the various meaning potentials: while one notices a few overlaps, one especially notices a good number of additions and more fine-grained descriptions as a result of the corpus analysis. The order of the meaning potentials that do overlap is not always the same either (cf. [ ] in Table 3).

What the manual effort does include, and what the corpus does not reveal in the same way, is the long list of 19 proverbs seen in Table 2. This is only partly the result of the fact that proverbs are known to be far less canonical in their use than dictionary-makers try to make you believe (Moon 1998). The proverb *Akaviile mu igi tikatya ikoli* 'The one that has just come from an egg does not fear an eagle' from the monolingual dictionary is for instance found in the corpus as *Akazaalibwa tikatya ikoli* 'The one which has just been born does not fear an eagle', hence without what one would assume to be a core term, 'egg'. Or, more Bantuish in nature, the monolingual-dictionary proverb *Omusaadha kikele kiva kyonka mu bwina* 'A man is a frog, which comes out of the hole by itself' is found in the corpus as *Omusaadha ikere: liva lyonka mu bwina* 'A man is a frog, which comes out of the hole by itself', which appears to be the same in translation, but in Lusoga the canonical form uses the noun in gender 7/8, while it is found in gender 5/6 in the corpus. Given this variation, proverbs have to be spotted mostly manually in a corpus. As to the reverse, a dedicated search does reveal proverbs not included into the otherwise pretty exhaustive manual list, such as *Awava omugulu waila mwigo* 'The stick takes the place of the leg that has left', *Awava omwosi wava omulilo* 'Fire comes from where smoke comes from', etc. Even so, their frequency of use is simply too low to merit inclusion when reasonable corpus frequencies and a nice spread across sources are used as an inclusion criterion.

3.6   Constructing corpus-driven microstructures for the Lusoga verb *-v-*

The data synthesised in Table 3 is the starting point for constructing the various dictionary articles that revolve around the verb *-v-* in Lusoga. In a desk or school dictionary, one may select from that data by taking, say, only the top *n* (frequent) lemmata and for these the top *n* (frequent) meaning potentials. At the other extreme, in a comprehensive dictionary, one will also want to exemplify all possible senses. To do so, reusing the examples that were studied during the analysis is an option, so the sentences and phrases from §§3.4.2–3.4.7 are prime candidates.[10] In doing so, however, it is good to recall that 'giving equal prominence to all senses, when they are not equally common, is a distortion' (Hanks 2002: 157). So, the most frequent meaning potentials could be illustrated with multiple examples, while the lesser-frequent ones could do with

just one or even no examples. Likewise with the combinations: whether or not to include some or all of them will depend on the target. For an unabridged paper dictionary, however, or for a digital dictionary in which the information is layered and where it may be 'peeled off' (Geeraerts 2000: 78-79), one can as well prepare and *optionally* present as much as possible. Adding the sources of the various examples also becomes a worthwhile addition at that point, as is the tradition in dictionaries based on historical principles. While such information on each source could be synthesised in the dictionary itself, a link to the full information, as seen in Addendum 1 of Part 1, could furthermore easily be added. In a digital dictionary actual hyperlinks to the corpus material itself could even be envisaged, thereby handing dictionary users the 'raw data' on which the lexicographers based their decisions, and/or allowing such users to explore the (corpus) data further (cf. de Schryver 2003: 167, 169, i.e. 'Dream # 31'). In short, a maximally populated dictionary writing system is best viewed as a *single* database from which *any number* of dictionaries may be generated, a concept that has been termed 'one database, many dictionaries' (de Schryver and Joffe 2005).

4.   Discussion

In this article we have made a strong case for the analysis of corpora to discover word meanings. After two decades of querying corpora for Bantu lexicography in general, and about one decade of corpus-building for Lusoga in particular, we are pretty much convinced that a careful study of the natural production of language that was produced by a multitude of speakers and writers indeed offers the best perspective on how language is truly used, from which meanings may be mapped (as explained and illustrated in the present article), and with which detailed studies of language may be undertaken. Some colleagues remain sceptical however, as voiced by Michael Marlo two years ago:

> A criticism that can be levelled at corpus-based approaches is that because they lump together data by individual speakers, it is extremely difficult if not impossible in a corpus-based approach to make sense of variation across individuals which is the result of the speakers having different internal grammars. The present approach seems to reject the idea that grammar is in the heads of individual speakers. It focuses on 'e-language' vs. 'i-language'. That is fine, but the approach has some limitations — such as the ability to state with precision what is a 'language'. (Marlo 2016, personal communication)

By using a corpus in the way we do, one ends up compromising, and indeed focusing on many e-languages (with *e* for 'external/externalised'), rather than on a single or a limited number of i-languages (with *i* for 'internal/internalised'). That said, even though the corpus analyst likes lots and lots of data and ditto examples, it is also true that: '"Overwhelming evidence", be it noted, may

consist of no more than a handful of textually well-formed and convincing modern uses' (Hanks 2002: 174). Michael Marlo goes on to suggest:

> Moreover, most linguists consider negative evidence to be essential for understanding the rules of language — not just what is common vs. uncommon but determining what is possible vs. impossible. There is considerable discussion of this within the generativist community under the notion of 'poverty of the stimulus' — the idea that speakers of a language know much about the language, even if they have never heard the expressions in question before. (Marlo 2016, personal communication)

In our strand of corpus linguistics, the focus is on the norms, not the exploitations, and the focus is consequently also not on what does not occur or on what occurs infrequently. Of course, this is a choice, but for a language like Lusoga which needs 'first descriptions', focusing on the speech community and their general needs first, and attempting to bring back their own words to them, in this case in the form of corpus-driven dictionary-making, seems like a worthwhile venture.

With this, we have come to the end of our three-part study of corpus-driven Bantu lexicography as applied to Lusoga. To conclude, it is now fitting to point out that our effort is not the first trilogy of articles on the application of corpora in modern dictionary-making. As a matter of fact, Michael Rundell and Penny Stock initiated this trend a quarter of a century ago, with a three-part report on what was then called 'The corpus revolution' (as applied to English lexicography). Compared to our effort, the sequence of their articles is organised differently, however. In their first part, Rundell and Stock (1992a) looked at the relative merits of large-scale text corpora compared to traditional citation banks. In the light of Hanks's theoretical framework of **mapping meaning onto use**, their most important observation in favour of the use of computerised corpora over manual reading and marking is that:

> It is astonishingly difficult for even the most experienced person to collect material for ordinary everyday usages since human beings tend to notice the unusual. [...] When using corpus evidence, therefore, the lexicographer works with whatever comes up in the corpus rather than with individually or specially selected examples. (Rundell and Stock 1992a: 13, 10)

The other advantages they list in favour of a corpus remain valid to this day, and have also all been illustrated for Lusoga lexicography: (i) 'it can provide evidence for the comparative frequency of word occurrence and behaviour', (ii) 'It can be of immense help in enabling the lexicographer to give examples to show the word in its most typically or frequently used contexts', (iii) 'It allows the lexicographer to structure an entry in such a way as to reflect how a word is normally used', and (iv) 'It can enable the dictionary maker to give an accurate account of grammatical behaviour at the level of individual senses' (Rundell and Stock 1992a: 14).

In their second part, Rundell and Stock (1992b) looked at the ways in which corpus evidence informs the actual writing of dictionary articles. With de Schryver and Joffe's practical concept of **one database, many dictionaries** in mind, the following observations on what to put in a certain dictionary ring true:

> In fact the task of omitting or not including known meanings which are none-theless inappropriate to a particular dictionary is a very hard one. It is so much easier to play safe and let such meanings in [...] Again the evidence of many millions of examples of usage can be of enormous assistance in strengthening the lexicographer's nerve in such cases [...] (Rundell and Stock 1992b: 25)

On a more generic level, their closing statement has proven to be as valid for Bantu as it is for English:

> It is perhaps fairly rare to find all one's preconceptions about a word being overturned on consulting a corpus, but it is equally rare to come away from analysing a given word or use without having learned a great deal that is new, illuminating, and sometimes unnerving. (Rundell and Stock 1992b: 28-29)

In their third part, Rundell and Stock (1992c) mainly deal with corpus building, and try to predict some of the automated tools and procedures that will be developed. These are, using the terms that have come to be adopted since Rundell and Stock's predictions from the early 1990s: (i) lemmatisers, (ii) sampling techniques, (iii) POS-taggers, (iv) parsers, and (v) word-sense disambiguators. Over the past 25 years these have indeed all been created for the world's major languages. More in particular, in Part 2 of our series we have indicated how the lemmatisation and POS-tagging for lexicographic purposes may be achieved for the Bantu languages. Unlike for English, these macrostructural aspects are hugely complex for the Bantu languages, which led Prinsloo and de Schryver to develop instruments known as **part-of-speech rulers and alphabetical (or multidimensional lexicographic) rulers** in order to measure, evaluate, predict and manage Bantu-language dictionary projects. We therefore trust that thanks to corpora, and just as is the case for English, we are now indeed 'emancipated from the role of harmless drudge and empowered to make new insights into every area of language' (Rundell and Stock 1992c: 51).

Abbreviations

#	noun class number	AP_x	adjectival prefix (of cl. x)
ADV	adverb	AUG	augment
AFP	affiliation prefix	CAUS	causative
APERF	after-perfect	cl.	class
APPL	applicative	CON	connective

COP	copulative	OM$_x$	object marker (of cl. or ps. x)
CP$_x$	connective prefix (of cl. x)	OREL$_x$	object relative (of cl. x)
DEM$_{A,B,C}$	demonstrative (of position A, B, C)	PASS	passive
		PERF	perfect
DEV	deverbative	PERS$_x$	personal pronoun (of ps. x)
ENCL$_x$	locative enclitic (x = cl. 16, 17, 18, 23)	PL	plural
		POSS$_x$	possessive (of ps. x)
EXC	exclusive pronoun	PP$_x$	pronominal pronoun (of cl. x)
ext.	extension		
FUT$_1$	future tense *-naa-*	PROG	progressive
FUT$_2$	future tense *-li-*	PRON$_x$	pronoun (of ps. x)
FV	final vowel	ps.	person
HAB	habitual	PST	past tense
INC	inclusive pronoun	REC	reciprocal
INTENS	intensifier	RFL	reflexive
INTER	interrogative	RPT	repetitive
INTERJ	interjection	SG	singular
loc.	locativised	SM$_x$	subject marker (of cl. or ps. x)
LOC$_x$	locative (x = cl. 16, 17, 18, 23)	SREL$_x$	subject relative (of cl. x)
MOD	modality	SUBJ	subjunctive
MODF	modifier	UNS	unspecified noun
NEG$_{A,B}$	negative (of type A, B)		

Acknowledgements

Endnotes

1.   Parts of this theoretical discussion are based on sections from Nabirye (2016).
2.   The reference to 'real' language is taken from the first-ever corpus-based dictionary, the *Collins COBUILD English Language Dictionary* (Sinclair 1987a), which was advertised as such.
3.   In corpus-linguistic circles, Sinclair may be best known as the founder of the COBUILD (i.e., the Collins Birmingham University International Language Database) project in lexical computing (Sinclair 1987b), and the chief editor of the *Collins COBUILD English Language Dictionary* (Sinclair 1987a). The managing editor of the latter dictionary was Patrick Hanks.
4.   Conversely, typologists may be interested in simply knowing what a language is capable of, and want answers to questions like: 'What is the longest possible verb form in this or that Bantu language?'
5.   Also present in the TLex database, but not frequent enough to have been lemmatised, are *-vw-*, a spoken variant of *-v-*, and *-evaamu* 'dare; be brave', the reflexive form of *-vaamu*.
6.   2 Timothy 2:21

7.   The locativised verb *-vaawo* also has a variant, namely *-vaagho*, but its frequency is too low to have made it into the lemmatised frequency list.

8.   ARVs = antiretrovirals (i.e., drugs to treat HIV)

9.   *Kodh'eyo* was a short-lived newspaper (1997–1998) written in Lusoga.

10. For the value of corpus examples over 'invented' (but more didactic) ones see Fox (1987).

References

Atkins, B.T.S. 1996. Bilingual Dictionaries: Past, Present and Future. Gellerstam, M., J. Järborg, S.-G. Malmgren, K. Norén, L. Rogström and C.R. Papmehl (Eds). 1996. *Euralex '96 Proceedings I–II, Papers Submitted to the Seventh EURALEX International Congress on Lexicography in Göteborg, Sweden*: 515-546. Gothenburg: Department of Swedish, Göteborg University.

Bosch, S.E. and D.J. Prinsloo. 2002. 'Abbreviated Nouns' in African Languages: A Morphological, Semantic and Lexicographic Perspective. *South African Journal of African Languages* 22(1): 92-104.

Bostoen, K. and G.-M. de Schryver. 2015. Linguistic Innovation, Political Centralization and Economic Integration in the Kongo Kingdom: Reconstructing the Spread of Prefix Reduction. *Diachronica* 32(2): 139-185 + 13 pages of supplementary material online.

Bostoen, K., F. Mberamihigo and G.-M. de Schryver. 2012. Grammaticalization and Subjectification in the Semantic Domain of Possibility in Kirundi (Bantu, JD62). *Africana Linguistica* 18: 5-40.

Chabata, E. and D. Nkomo. 2010. The Utilisation of Outer Texts in the Practical Lexicography of African Languages. *Lexikos* 20: 73-91.

De Kind, J. and K. Bostoen. 2012. The Applicative in ciLubà Grammar and Discourse: A Semantic Goal Analysis. *Southern African Linguistics and Applied Language Studies* 30(1): 101-124.

De Kind, J., M. Devos, G.-M. de Schryver and K. Bostoen. 2013. Negation Markers, Focus Markers and Jespersen Cycles in Kikongo (Bantu, H16): A Comparative and Diachronic Corpus-based Approach. Available online at: https://www2.hu-berlin.de/predicate_focus_africa/data/2013-12-10_deKind_Negation.in.Kikongo.pdf.

De Kind, J., S. Dom, G.-M. de Schryver and K. Bostoen. 2015. Event-centrality and the Pragmatics-Semantics Interface in Kikongo: From Predication Focus to Progressive Aspect and Vice Versa. *Folia Linguistica Historica* 36: 113-163.

de Schryver, G.-M. 1999. *Cilubà Phonetics, Proposals for a 'Corpus-based Phonetics from Below'-Approach* (Recall Linguistics Series 14). Ghent: Recall.

de Schryver, G.-M. 2002. Web for/as Corpus: A Perspective for the African Languages. *Nordic Journal of African Studies* 11(2): 266-282.

de Schryver, G.-M. 2003. Lexicographers' Dreams in the Electronic-Dictionary Age. *International Journal of Lexicography* 16(2): 143-199.

de Schryver, G.-M. 2005. Book Review: M.-H. Corréard, ed. 2002. Lexicography and Natural Language Processing. A Festschrift in Honour of B.T.S. Atkins. *Lexicographica: International Annual for Lexicography* 21: 420-425.

de Schryver, G.-M. 2006. Compiling Modern Bilingual Dictionaries for Bantu Languages: Case Studies for Northern Sotho and Zulu. Corino, E., C. Marello and C. Onesti (Eds). 2006. *Atti del XII Congresso Internazionale di Lessicografia, Torino, 6-9 settembre 2006 / Proceedings XII Euralex International Congress, Torino, Italia, September 6th–9th, 2006*: 515-525. Alessandria: Edizioni dell'Orso.

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: Oxford University Press Southern Africa.

de Schryver, G.-M. 2008. Why does Africa need Sinclair? *International Journal of Lexicography* 21(3): 267-291.

de Schryver, G.-M. and R. Gauton. 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. and D. Joffe. 2005. One Database, Many Dictionaries — Varying Co(n)text with the Dictionary Application TshwaneLex. Ooi, V.B.Y., A. Pakir, I. Talib, L. Tan, P.K.W. Tan and Y.Y. Tan (Eds). 2005. *Words in Asian Cultural Contexts, Proceedings of the 4th Asialex Conference, 1–3 June 2005, M Hotel, Singapore*: 54-59. Singapore: Department of English Language and Literature & Asia Research Institute, National University of Singapore.

de Schryver, G.-M. and B. Lepota. 2001. The Lexicographic Treatment of Days in Sepedi, or When Mother-Tongue Intuition Fails. *Lexikos* 11: 1-37.

de Schryver, G.-M. and M. Nabirye. 2010. A Quantitative Analysis of the Morphology, Morpho-phonology and Semantic Import of the Lusoga Noun. *Africana Linguistica* 16: 97-153.

de Schryver, G.-M. and D.J. Prinsloo. 2000. 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.

de Schryver, G.-M. and D.J. Prinsloo. 2001. Towards a Sound Lemmatisation Strategy for the Bantu Verb through the Use of Frequency-based Tail Slots — with Special Reference to Cilubà, Sepedi and Kiswahili. Mdee, J.S. and H.J.M. Mwansoko (Eds). 2001. *Makala ya kongamano la kimataifa Kiswahili 2000. Proceedings*: 216-242, 372. Dar es Salaam: TUKI, Chuo Kikuu cha Dar es Salaam.

de Schryver, G.-M. and E. Taljard. 2006. Locative Trigrams in Northern Sotho, Preceded by Analyses of Formative Bigrams. *Linguistics, An Interdisciplinary Journal of the Language Sciences* 44(1): 135–193.

de Schryver, G.-M. and E. Taljard. 2007. Compiling a Corpus-based Dictionary Grammar: An Example for Northern Sotho. *Lexikos* 17: 37-55.

de Schryver, G.-M., E. Taljard, M.P. Mogodi and S. Maepa. 2004. The Lexicographic Treatment of the Demonstrative Copulative in Sesotho sa Leboa — An Exercise in Multiple Cross-referencing. *Lexikos* 14: 35-66.

Devos, M. and G.-M. de Schryver. 2013. From 'habitually going' to 'maybe': Grammaticalization and Lexicalization of an Epistemic Sentence Adverb in Swahili. *Abstracts of The 21st International Conference on Historical Linguistics*: 29. Oslo: University of Oslo.

Devos, M. and G.-M. de Schryver. 2016. From Usually Going to Epistemic Possibility. Origin and Development of an Epistemic Sentence Adverb in Swahili. *6th International Conference on Bantu Languages, Workshop on the Expression of Mood and Modality in Bantu Languages*: 11. Helsinki: University of Helsinki.

Devos, M., M.-J. Misago and K. Bostoen. 2017. A Corpus-based Description of Locative and Non-locative Reference in Kirundi Locative Enclitics. *Africana Linguistica* 23: 47-83.

Dom, S., G. Segerer and K. Bostoen. 2015. Antipassive/Associative Polysemy in Cilubà (Bantu, L31a): A Plurality of Relations Analysis. *Studies in Language* 39(2): 354-385.

Faaß, G. 2010. *A Morphosyntacic Description of Northern Sotho as a Basis for an Automated Translation from Northern Sotho into English*. Unpublished Ph.D. dissertation. Pretoria: University of Pretoria.

Fox, G. 1987. The Case for Examples. Sinclair, J.M. (Ed.). 1987. *Looking Up: An Account of the COBUILD Project in Lexical Computing and the Development of the Collins COBUILD English Language Dictionary*: 137-149. London: Collins ELT.

Gauton, R., G.-M. de Schryver and L. Mohlala. 2004. A Corpus-based Investigation of the Zulu Nominal Suffix -kazi: A Preliminary Study. Akinlabi, A. and O. Adesola (Eds). 2004. *Proceedings of the 4th World Congress of African Linguistics, New Brunswick 2003*: 373-380. Cologne: Rüdiger Köppe Verlag.

Geeraerts, D. 2000. Adding Electronic Value. The Electronic Version of the *Grote Van Dale*. Heid, U., S. Evert, E. Lehmann and C. Rohrer (Eds). 2000. *Proceedings of the Ninth EURALEX International Congress, EURALEX 2000, Stuttgart, Germany, August 8th–12th, 2000*: 75-84. Stuttgart: Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.

Geeraerts, D. 2009. Currents and Undercurrents in Lexical Semantics, Twenty Years After. Beijk, E., L. Colman, M. Göbel, F. Heyvaert, T. Schoonheim, R. Tempelaars and V. Waszink (Eds). 2009. *Fons Verborum. Feestbundel voor prof. dr. A.M.F.J. (Fons) Moerdijk, aangeboden door vrienden en collega's bij zijn afscheid van het Instituut voor Nederlandse Lexicologie*: 421-430. Amsterdam: Gopher BV.

Geeraerts, D. 2010. *Theories of Lexical Semantics*. New York: Oxford University Press.

Gouws, R.H. and D.J. Prinsloo. 1997. Lemmatisation of Adjectives in Sepedi. *Lexikos* 7: 45-57.

Gouws, R.H. and D.J. Prinsloo. 2005. Left-expanded Article Structures in Bantu with Special Reference to isiZulu and Sepedi. *International Journal of Lexicography* 18(1): 25-46.

Hanks, P. 2002. Mapping Meaning onto Use. Corréard, M.-H. (Ed.). 2002. *Lexicography and Natural Language Processing. A Festschrift in Honour of B.T.S. Atkins*: 156-198. s.l.: Euralex.

Hanks, P. 2004. Corpus Pattern Analysis. Williams, G. and S. Vessier (Eds). 2004. *Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004, Lorient, France, July 6–10, 2004*: 87-97. Lorient: Faculté des Lettres et des Sciences Humaines, Université de Bretagne Sud.

Hanks, P. 2012. The Corpus Revolution in Lexicography. *International Journal of Lexicography* 25(4): 398-436.

Hanks, P. 2013. *Lexical Analysis: Norms and Exploitations*. Cambridge, MA: The MIT Press.

Hanks, P. 2015. Cognitive Semantics and the Lexicon. *International Journal of Lexicography* 28(1): 86-106.

Joffe, D. and G.-M. de Schryver. 2002–18. TLex Suite — Dictionary Compilation Software. Available online at: http://tshwanedje.com/tshwanelex/.

Kawalya, D. 2017. *A Corpus-driven Study of the Expression of Modality in Luganda (Bantu, JE15)*. Unpublished Ph.D. dissertation. Ghent: Ghent University.

Kawalya, D., K. Bostoen and G.-M. de Schryver. 2014. Diachronic Semantics of the Modal Verb -sóból- in Luganda: A Corpus-driven Approach. *International Journal of Corpus Linguistics* 19(1): 60-93.

Kawalya, D., G.-M. de Schryver and K. Bostoen. 2018. From Conditionality to Modality in Luganda (Bantu, JE15): A Synchronic and Diachronic Corpus Analysis of the Verbal Prefix -andi-. *Journal of Pragmatics* 127: 84-106.

Kilgarriff, A. 2003–18. Sketch Engine. Available online at: https://www.sketchengine.co.uk.

Klein, J. 2010a. Can the New African Language Dictionaries Empower the African Language Speakers of South Africa or Are They Just a Half-hearted Implementation of Language Policies? Dykstra, A. and T. Schoonheim (Eds). 2010. *Proceedings of the XIV Euralex International Congress (Leeuwarden, 6–10 July 2010)*: 1485-1496. Leeuwarden/Ljouwert: Fryske Akademy.

Klein, J. 2010b. Nord Sotho Wörterbücher als Implementierungsstrategien der *South African Languages Bill* und der *National Lexicographic Units Bill*. Buchmann, L., L. Fuhrmann, N. Nassenstein, C. Vogel, M. Weinle and A. Wolvers (Eds). 2010. *Beiträge zur 3. Kölner Afrikawissenschaftlichen Nachwuchstagung (KANT III)*: 1-11. Cologne: University of Cologne.

Kosch, I. 2013. An Analysis of the *Oxford Bilingual School Dictionary: Northern Sotho and English* (De Schryver 2007). *Lexikos* 23: 611-627.

Lafkioui, M., E. Nshemezimana and K. Bostoen. 2016. Cleft Constructions and Focus in Kirundi. *Africana Linguistica* 22: 71-106.

Madiba, M. and D. Nkomo. 2010. The *Tshivenḓa-English Ṱhalusamaipfi/Dictionary* as a Product of South African Lexicographic Processes. *Lexikos* 20: 307-325.

Mberamihigo, F. 2014. *L'expression de la modalité en kirundi : Exploitation d'un corpus électronique*. Unpublished Ph.D. dissertation. Brussels; Ghent: Université libre de Bruxelles; Ghent University.

Mberamihigo, F., G.-M. de Schryver and K. Bostoen. 2016. Entre verbe et adverbe : Grammaticalisation et dégrammaticalisation du marqueur épistémique *umeengo/umeenga* en kirundi (bantou, JD62). *Journal of African Languages and Linguistics* 37(2): 247-286.

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

Moon, R. 1998. *Fixed Expressions and Idioms in English. A Corpus-based Approach*. Oxford: Oxford University Press.

Nabirye, M. 2008. *Compilation of the Monolingual Lusoga Dictionary*. Unpublished M.A. dissertation. Kampala: Makerere University.

Nabirye, M. 2009a. Compiling the First Monolingual Lusoga Dictionary. *Lexikos* 19: 177-196.

Nabirye, M. 2009b. *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]*. Kampala: Menha Publishers.

Nabirye, M. 2016. *A Corpus-based Grammar of Lusoga*. Unpublished Ph.D. dissertation. Ghent: Ghent University.

Nabirye, M. and G.-M. de Schryver. 2010. The Monolingual Lusoga Dictionary Faced with Demands from a New User Category. *Lexikos* 20: 326-350.

Nabirye, M. and G.-M. de Schryver. 2011. From Corpus to Dictionary: A Hybrid Prescriptive, Descriptive and Proscriptive Undertaking. *Lexikos* 21: 120-143.

Nabirye, M. and G.-M. de Schryver. 2013. Digitizing the Monolingual Lusoga Dictionary: Challenges and Prospects. *Lexikos* 23: 297-322.

Nong, S., G.-M. de Schryver and D.J. Prinsloo. 2002. Loan Words versus Indigenous Words in Northern Sotho — A Lexicographic Perspective. *Lexikos* 12: 1-20.

Nshemezimana, E. 2016. *Morphosyntaxe et structure informationnelle en kirundi : Focus et stratégies de focalisation*. Unpublished Ph.D. dissertation. Ghent: Ghent University.

Nshemezimana, E. and K. Bostoen. 2016. The Conjoint/Disjoint Alternation in Kirundi (JD62): A Case for its Abolition. van der Wal, J. and L.M. Hyman (Eds). 2016. *The Conjoint/Disjoint*

Alternation in Bantu (Trends in Linguistics. Studies and Monographs 301): 390-425. Berlin: Mouton de Gruyter.

Prinsloo, D.J. 1992. Lemmatization of Reflexives in Northern Sotho. *Lexikos* 2: 178-191.

Prinsloo, D.J. 1994. Lemmatization of Verbs in Northern Sotho. *South African Journal of African Languages* 14(2): 93-102.

Prinsloo, D.J. 2002. The Lemmatization of Copulatives in Northern Sotho. *Lexikos* 12: 21-43.

Prinsloo, D.J. 2003. The Lemmatisation of Adverbs in Northern Sotho. *Lexikos* 13: 21-37.

Prinsloo, D.J. 2005. Electronic Dictionaries Viewed from South Africa. *Hermes, Journal of Linguistics* 34: 11-35.

Prinsloo, D.J. 2006. Compiling a Bidirectional Dictionary Bridging English and the Sotho Languages: A Viability Study. *Lexikos* 16: 193-204.

Prinsloo, D.J. 2009. Current Lexicography Practice in Bantu with Specific Reference to the *Oxford Northern Sotho School Dictionary*. *International Journal of Lexicography* 22(2): 151-178.

Prinsloo, D.J. 2012. Die leksikografiese bewerking van verwantskapsterme in Sepedi. *Lexikos* 22: 272-289.

Prinsloo, D.J. 2014a. A Critical Evaluation of the Paradigm Approach in Sepedi Lemmatisation — The *Groot Noord-Sotho Woordeboek* as a Case in Point. *Lexikos* 24: 251-271.

Prinsloo, D.J. 2014b. Lexicographic Treatment of Kinship Terms in an English / Sepedi–Setswana–Sesotho Dictionary with an Amalgamated Lemmalist. *Lexikos* 24: 272-290.

Prinsloo, D.J. and S.E. Bosch. 2012. Kinship Terminology in English–Zulu / Northern Sotho Dictionaries — A Challenge for the Bantu Lexicographer. Fjeld, R.V. and J.M. Torjusen (Eds). 2012. *Proceedings of the 15th EURALEX International Congress, 7–11 August, 2012, Oslo*: 296-303. Oslo: Department of Linguistics and Scandinavian Studies, University of Oslo.

Prinsloo, D.J., T.J.D. Bothma and U. Heid. 2014. User Support in e-Dictionaries for Complex Grammatical Structures in the Bantu Languages. Abel, A., C. Vettori and N. Ralli (Eds). 2014. *Proceedings of the XVI EURALEX International Congress: The User in Focus, 15–19 July 2014, Bolzano/Bozen*: 819-827. Bolzano/Bozen: EURAC Research.

Prinsloo, D.J., T.J.D. Bothma, U. Heid and D.J. Prinsloo. 2017. Direct User Guidance in e-Dictionaries for Text Production and Text Reception — The Verbal Relative in Sepedi as a Case Study. *Lexikos* 27: 403-426.

Prinsloo, D.J. and G.-M. de Schryver. 1999. The Lemmatization of Nouns in African Languages with Special Reference to Sepedi and Cilubà. *South African Journal of African Languages* 19(4): 258-275.

Prinsloo, D.J. and G.-M. de Schryver. 2001. Taking Dictionaries for Bantu Languages into the New Millennium — with Special Reference to Kiswahili, Sepedi and isiZulu. Mdee, J.S. and H.J.M. Mwansoko (Eds). 2001. *Makala ya kongamano la kimataifa Kiswahili 2000. Proceedings*: 188-215. Dar es Salaam: TUKI, Chuo Kikuu cha Dar es Salaam.

Prinsloo, D.J. and G.-M. de Schryver. 2002. Reversing an African-language Lexicon: The *Northern Sotho Terminology and Orthography No. 4* as a Case in Point. *South African Journal of African Languages* 22(2): 161-185.

Prinsloo, D.J. and R.H. Gouws. 1996. Formulating a New Dictionary Convention for the Lemmatization of Verbs in Northern Sotho. *South African Journal of African Languages* 16(3): 100-107.

Prinsloo, D.J. and R.H. Gouws. 2006. Lexicographic Presentation of Grammatical Divergence in Sesotho sa Leboa. *South African Journal of African Languages* 26(4): 184-197.

Prinsloo, D.J., U. Heid, T.J.D. Bothma and G. Faaß. 2012. Devices for Information Presentation in Electronic Dictionaries. *Lexikos* 22: 290-320.

Rundell, M. and P. Stock. 1992a. The Corpus Revolution 1. The first in a series of three reports on the development and use of electronic language corpora and their impact on dictionaries. *English Today, The International Review of the English Language* 8(2): 9-14.

Rundell, M. and P. Stock. 1992b. The Corpus Revolution 2. A consideration of the practical benefits to English-language lexicographers of the evidence derived from computer corpora (second article of three). *English Today, The International Review of the English Language* 8(3): 21-29.

Rundell, M. and P. Stock. 1992c. The Corpus Revolution 3. A consideration of the prospects and potential of corpus-and-concordance lexicography (third article of three). *English Today, The International Review of the English Language* 8(4): 45-51.

Scannell, K.P. 2003–18. An Crúbadán — Corpus Building for Minority Languages. Available online at: http://crubadan.org/.

Scott, M. 1996–2018. WordSmith Tools. Available online at: http://www.lexically.net/wordsmith/.

Sewangi, S.S. 2000. Tapping the Neglected Resource in Kiswahili Terminology: Automatic Compilation of the Domain-specific Terms from Corpus. *Nordic Journal of African Studies* 9(2): 60-84.

Sewangi, S.S. 2001. *Computer-assisted Extraction of Terms in Specific Domains: The Case of Swahili*. Unpublished PhD dissertation. Helsinki: University of Helsinki.

Sinclair, J.M. 1966. Beginning the Study of Lexis. Bazell, C.E., J.C. Catford, M.A.K. Halliday and R.H. Robins (Eds). 1966. *In Memory of J.R. Firth*: 410-430. London: Longmans.

Sinclair, J.M. 1987a. *Collins COBUILD English Language Dictionary*. London: William Collins Sons & Co.

Sinclair, J.M. (Ed.). 1987b. *Looking Up: An Account of the COBUILD Project in Lexical Computing and the Development of the Collins COBUILD English Language Dictionary*. London: Collins ELT.

Taljard, E. 2006. Corpus-based Linguistic Investigation for the South African Bantu Languages: A Northern Sotho Case Study. *South African Journal of African Languages* 26(4): 165-183.

Taljard, E. 2012. Corpus-based Language Teaching: An African Language Perspective. *Southern African Linguistics and Applied Language Studies* 30(3): 377-393.

Taljard, E. and G.-M. de Schryver. 2002. Semi-automatic Term Extraction for the African Languages, with Special Reference to Northern Sotho. *Lexikos* 12: 44-74.

Taljard, E. and G.-M. de Schryver. 2016. A Corpus-driven Account of the Noun Classes and Genders in Northern Sotho. *Southern African Linguistics and Applied Language Studies* 34(2): 169-185.

Tognini-Bonelli, E. 2001. *Corpus Linguistics at Work* (Studies in Corpus Linguistics 6). Amsterdam: John Benjamins.

Toscano, M. and S.S. Sewangi. 2005. Discovering Usage Patterns for the Swahili *amba-* Relative Forms cl. 16, 17, 18: Using Corpus Data to Support Autonomous Learning of Kiswahili by Italian Speakers. *Nordic Journal of African Studies* 14(3): 274-317.

Tummers, J., K. Heylen and D. Geeraerts. 2005. Usage-based Approaches in Cognitive Linguistics: A Technical State of the Art. *Corpus Linguistics and Linguistic Theory* 1(2): 225-261.

Addendum 1:   *Eiwanika ly'Olusoga* (Nabirye 2009b), page 379

(o)ku.v.a

(o)ku.v.a[1] [(o)kúvá] *kt.* **[L]** *[-viile]* [nviile] **bl:** [**Lg:** okuva] 1) Okusimbuka mu kifo ekilala waayolekela ekindi. *gez: Nva Mayuge.* 2) Okusibuka. *gez: Nva Iganga.* 3) Okulekelela ekintu ky'obaile okola. *gez: Ebyo nabiviileku.* 4) Okuseguka mu ngila oba mu kifo. *gez: Leka nkuviile ofune eidembe.* 'sk: Okuva ku luguudo: Okwonooneka / Okuva ku mulembe.'. **ggl:** Awava akwita n'awava akukobela; Awava ennume waila nnume. Awava mwino tiwaila mwino: awava eliino waila ilibu; Awava mwino tiwaila mwino: Awava eliiso waila itulu; Akaviile mu igi tikatya ikoli; Edhiva okulala n'embilo; Empambo eva ku kiwalo; Ennhonhi eva ewala temala mutonto; Ensanafu eva ku mugendelo telwa kufuuka kabasa; Atava ku mulungi afa t'awoza; Ka nduviile ku ntobo oti n'omuyala atuuse we bafumba; Olusubi olulala we luva ku ndhu tetoonha; Omukazi omulungi nnimilo ya ngila buli avaayo agyegwaniza; Omukwano guva mu ngabo; Omukwano guva mu ngila gwatuuka eka; Omusaadha kikele kiva kyonka mu bwina; Va we ndi takulwania; Va ku ntebe ya lata awulilila ku ise; W'ova tosoile w'otela okwila.
bbgz: Okuviila, Okuviisa.

(o)ku.v.a[2] [(o)kúvá] *kt.* **[L]** *[-viile]* [nviile] **bl:** [**Lg:** okuva] Okutandiikila mu kifo ekilala okutuuka ku kifo ekindi. *gez: Ennhandha ya Nalubaale eva Indhindha.*

(o)ku.vaabiil.a [(o)kúváábíílá] *kt.* **[L]** *[-vaabiile]* [nvaabiile] **bl:** [**Lg:** okuvaabiila] 1) Okuvuukila ebyokulya. 2) Okuliisa amailu.
bbgz: Okuvaabiilila, Okuvaabiiza.

(o)ku.vaal.a [(o)kúvaalá] *kt.* **[L]** *[-vaile]* [nvaile] **bbl:** [**Lsw:** vaa, **Lg:** okuvaala] 1) Okubisa omutwe, emikono, amagulu oba ebigele mu kintu ng'olugoye oba engaito. *gez: Abantu bavaala engoye baleke kubita madu.*
bbgz: *Okuvaalibwa, Okuvaalika, Okuvaalila, Okuvaaza.*

(o)ku.vool.a

(o)ku.v.a.a.ku [(o)kúvááku] *kt.* **[T]** *[-viileku]* [nviileku] **bl:** [**Lg:** okuvaako] 1) Okuzima. *gez: Amasaanhalaze gaviileku.* 2) Okuleka. *gez: Omwenge nguviileku. / Oyo namuvaaku naafuna owundi.*
bbgz: *Okuviilwaku, Okuviisaaku.*

(o)ku.v.a.a.mu [(o)kúváámu] *kt.* **[T]** *[-vilemu]* [nviilemu] **bl:** [**Lg:** okuvaamu] 1) Obutatuukiliza kye wasuubiza omuntu. *gez: Tubaile tusuubiila nti agya kutuyamba aye atuviilemu.* 2) Okulyamu olukwe. *gez: Gwetwateesa naye mwene neeyatuvaamu.* 3) Okuwa oba okumaliliza. *gez: Bw'obifumba bivaamu bulungi.* 4) Okubula ky'ofuna oba ky'ogobolola mu kintu. *gez: Nze nviiliilemu awo.* 5) Okusagala. *gez: Eno engaito evaamu.*
bbgz: *Okuviilamu, Okuviisamu.*

(o)ku.vangan.a [(o)kúvangáná] *kt.* **[T]** *[-vangaine]* [nvangaine] 1) Okwagaanana okw'abantu abava mu bifo eby'endhawulo baawaya. 2) Okusonhiwagana mwailagana nga mubaile muyombye mwayawukana. 3) Okugaitagaita.
bbgz: *Okuvangania, Okuvanganila.*

(o)ku.vook.a [(o)kúvooká] **tbk:** Okúvootóká *kt.* **[L]** *[-voose]* Omubuli gw'ekintu okwenhiga gwaila munda oba gwefunha. *gez: Enthupa evoose.*
bbgz: *Okuvookela, Okuvoosa.*

(o)ku.vool.a[1] [(o)kúvóólá] *kt.* **[L]** *[-voile]* [nvoile] Okunhigiliza omubili gw'ekintu gwaingila munda ave nga tikimenheike. *gez: Oniinhie ku mukebe waaguvoola.*
bbgz: *Okuvoolebwa, Okuvooleka, Okuvooza, Okwevoola.*

(o)ku.vool.a[2] [(o)kúvoolá] *kt.* **bl:** [**Lg:** okuvvoola] *(stl)* Okunhooma oba okugaya omuwendo gw'ekintu. *gez: Ekigambo ekyo kivoola embeela y'omuntu.*

Addendum 2: Top orthographic corpus types in the 1.7m Lusoga corpus underlying the verbal lemmas *-v-*, *-vaawo* (cl. 16), *-vaaku* (cl. 17), *-vaamu* (cl. 18) and *-vaayo* (cl. 23) as well as the nominal *-vo* in cl. 14

Types in the 1.7m Lusoga corpus considered for the lemma *-v-*, verb:

abaava [-va] *verb, transitive cl. 2* **1** came from; **2** diverted / *freq.70* rank 2606 # texts 22;

abava [-va] *verb, transitive cl. 2* come from / *freq.81* rank 2289 # texts 32;

abaviile [-va] *verb, transitive cl. 2* **1** have come from; **2** have left / *freq.13* rank 10420 # texts 6;

agava [-va] *verb, transitive cl. 6* come from / *freq.43* rank 3953 # texts 21;

aghava [-va] *verb, transitive cl. 16* **1** comes from; **2** leaves / *freq.22* rank 6788 # texts 13;

ava [-va] *verb, transitive cl. 1* from / *freq.389* rank 509 # texts 107;

ave [-va] *verb, intransitive cl. 1* **1** comes from; **2** leaves / *freq.37* rank 4478 # texts 17;

aviile [-va] *verb, transitive cl. 1* **1** came from; **2** diverted; **3** left / *freq.35* rank 4655 # texts 25;

aviire [-va] *verb, transitive cl. 1* **1** has come from; **2** has left / *freq.13* rank 10467 # texts 8;

awava [-va] *verb, transitive cl. 16* **1** comes from; **2** departs / *freq.13* rank 10468 # texts 5;

baava [-va] *verb, transitive cl. 2* **1** came from; **2** diverted / *freq.118* rank 1598 # texts 40;

bava [-va] *verb, transitive cl. 2* from / *freq.120* rank 1572 # texts 47;

baviile [-va] *verb, transitive cl. 2* **1** have come from; **2** have left / *freq.13* rank 10514 # texts 10;

baviire [-va] / *freq.17* rank 8404 # texts 10 {Notes: see baviile};

biva [-va] *verb, transitive cl. 8* **1** from; **2** come from / *freq.63* rank 2851 # texts 35;

biviire [-va] *verb, transitive cl. 8* **1** came from; **2** diverted / *freq.104* rank 1787 # texts 7;

buva [-va] *verb, transitive cl. 14* come from / *freq.53* rank 3296 # texts 35;

byava [-va] *verb, transitive cl. 8* came from / *freq.14* rank 9926 # texts 8;

dhaava [-va] *verb, transitive cl. 10* came from / *freq.15* rank 9399 # texts 10;

dhiva [-va] *verb, transitive cl. 9* come from / *freq.26* rank 5968 # texts 19;

ebiva [-va] *verb, transitive cl. 8* come from / *freq.106* rank 1753 # texts 35;

ebyava [-va] *verb, transitive cl. 2* **1** came from; **2** resulted from / *freq.30* rank 5300 # texts 18;

edhiva [-va] *verb, transitive cl. 10* come from / *freq.29* rank 5457 # texts 18;

ekiva [-va] *verb, transitive cl. 7* comes from / *freq.38* rank 4408 # texts 16;

ekyava [-va] *verb, transitive cl. 7* **1** came from; **2** left / *freq.18* rank 8060 # texts 7;

eva [-va] *verb, transitive cl. 9* from / *freq.118* rank 1602 # texts 44;

eyava [-va] *verb, transitive cl. 1* came from / *freq.46* rank 3748 # texts 21;

gava [-va] *verb, transitive cl. 6* come from / *freq.33* rank 4920 # texts 25;

guva [-va] *verb, transitive cl. 3* comes from / *freq.27* rank 5809 # texts 16;

gwava [-va] *verb, transitive cl. 3* **1** came from; **2** left / *freq.13* rank 10656 # texts 11;

kava [-va] *verb, transitive cl. 12* **1** comes from; **2** departs / *freq.13* rank 10693 # texts 9;

kikuviire [-va] *verb, transitive cl. 7 + 15* **1** has escaped; **2** has come from / *freq.17* rank 8539 # texts 1;

kiva [-va] *verb, transitive cl. 7* from / *freq.94* rank 1984 # texts 39;

kuva [-va] *verb, transitive* **1** leave; **2** let alone / *freq.267* rank 749 # texts 96;

kwava [-va] *verb, transitive* **1** came from; **2** resulted from / *freq.20* rank 7437 # texts 4;

kyava [-va] *verb, transitive cl. 7* **1** came from; **2** diverted / *freq.51* rank 3434 # texts 24 {Notes: see also kye & ava > ky'ava};

kyebaava [-va] / *freq.19* rank 7772 # texts 5 {Notes: see kye & baava};

kyenva [-va] / *freq.36* rank 4597 # texts 17 {Notes: see kye & nva};

kyetuva [-va] / *freq.13* rank 10766 # texts 10 {Notes: see kye & tuva};

kyeyava [-va] *verb, transitive 7 + 1* that is why / *freq.143* rank 1329 # texts 16 {Notes: see kye & yava};

liva [-va] *verb, transitive cl. 11* comes from / *freq.32* rank 5052 # texts 15;

luva [-va] *verb, transitive cl. 11* **1** comes from; **2** results from / *freq.20* rank 7443 # texts 12;

lyava [-va] *verb, transitive* came from / *freq.17* rank 8589 # texts 6;

muva [-va] *verb, transitive 2pl.* come from / *freq.30* rank 5351 # texts 15;

naava [-va] *verb, transitive 1sg.* **1** left; **2** diverted; **3** parted with / *freq.20* rank 7476 # texts 13;

nava [-va] *verb, transitive 1sg.* **1** came from; **2** left / *freq.69* rank 2654 # texts 27;

nva [-va] *verb, transitive 1sg.* from / *freq.162* rank 1190 # texts 71;
nviile [-va] *verb, transitive 1sg.* **1** have left; **2** have come from / *freq.18* rank 8200 # texts 16;
nviire [-va] / *freq.17* rank 8657 # texts 8 {Notes: see nviile};
obuva [-va] *verb, transitive* come from / *freq.32* rank 5086 # texts 17;
oguva [-va] *verb, transitive cl. 3* comes from / *freq.26* rank 6067 # texts 11;
okuva [-va] *verb, transitive* **1** since; **2** from; **3** from ... to; **4** to leave or be from / *freq.2668* rank 58 # texts 207;
okuviila [-va] *verb, transitive* **1** leave; **2** come from / *freq.27* rank 5887 # texts 18;
okuviira [-va] *verb, transitive* **1** come from; **2** move away from / *freq.58* rank 3075 # texts 18;
oluva [-va] *verb, transitive cl. 11* **1** from which; **2** comes from / *freq.40* rank 4260 # texts 28;
olwava [-va] *verb, auxiliary cl. 11* came from / *freq.30* rank 5382 # texts 12;
omutuviira [-va] *cl. 1 + 1pl.* leave / *freq.93* rank 2015 # texts 5;
omuva [-va] *verb, transitive cl. 3* **1** from which; **2** comes from / *freq.42* rank 4090 # texts 28;
ova [-va] *verb, transitive 2sg.* **1** return; **2** reason / *freq.325* rank 619 # texts 103;
tava [-va] *verb, transitive 2sg.* does not come from / *freq.19* rank 7919 # texts 16;
tova [-va] *verb, transitive* **1** leave not; **2** divert not / *freq.15* rank 9729 # texts 13;
tunaava [-va] *verb, transitive 1pl. + 1sg.* **1** will leave; **2** will come from / *freq.14* rank 10340 # texts 9;
tuviile [-va] *verb, transitive 1pl.* **1** came from; **2** left / *freq.44* rank 3941 # texts 21;
twava [-va] *verb, transitive 1pl.* **1** came from; **2** diverted; **3** left / *freq.38* rank 4465 # texts 23;
va [-va] *verb, transitive* **1** leave; **2** let alone / *freq.81* rank 2312 # texts 42;
yaava [-va] *verb, transitive cl. 1, cl. 9* **1** come from; **2** diverted / *freq.96* rank 1955 # texts 45;
yava [-va] *verb, transitive cl. 1* from / *freq.188* rank 1041 # texts 55

Types in the 1.7m Lusoga corpus considered for the lemma *-vaawo*, loc. verb:

okuvaawo [-va-wo] *verb, intransitive* **1** leave; **2** become extinct / *freq.15* rank 9671 # texts 10

Types in the 1.7m Lusoga corpus considered for the lemma *-vaaku*, loc. verb:

kuvaaku [-va-ku] *verb, transitive* **1** cause; **2** trigger / *freq.14* rank 10102 # texts 14

Types in the 1.7m Lusoga corpus considered for the lemma *-vaamu*, loc. verb:

avaamu [-va-mu] *verb, transitive cl. 1.* **1** departs; **2** results into; **3** releases / *freq.18* rank 7990 # texts 15
bituviiremu [-va-mu] *verb, intransitive cl. 8 + 1pl.* **1** result into; **2** have come from / *freq.13* rank 10527 # texts 5
ebivaamu [-va-mu] *verb, transitive cl. 8* **1** come out; **2** become visible / *freq.12* rank 11353 # texts 12
ekivaamu [-va-mu] *verb, transitive cl. 7* comes out finally / *freq.37* rank 4494 # texts 24
ekyavaamu [-va-mu] *verb, transitive cl. 7* **1** came out; **2** resulted into / *freq.78* rank 2386 # texts 21
kuvaamu [-va-mu] **1** get out; **2** come forward; **3** betray; **4** end up / *freq.24* rank 6442 # texts 17
mwavaamu [-va-mu] *verb, transitive cl .18* **1** resulted into; **2** came out / *freq.14* rank 10180 # texts 9
okuvaamu [-va-mu] *verb, transitive* **1** leave; **2** get out; **3** end up / *freq.46* rank 3775 # texts 30

Types in the 1.7m Lusoga corpus considered for the lemma *-vaayo*, loc. verb:

avaayo [-va-yo] *verb, intransitive cl. 1.* **1** comes out; **2** surfaces / *freq.35* rank 4654 # texts 25
aveeyo [-va-yo] *verb, intransitive cl. 1.* **1** move away; **2** become visible / *freq.14* rank 9851 # texts 13
baveeyo [-va-yo] *verb, intransitive cl. 2* **1** become visible; **2** come out; **3** move away / *freq.13* rank 10513 # texts 12
kuvaayo [-va-yo] *verb, intransitive* **1** come out; **2** become visible; **3** move away / *freq.38* rank 4424 # texts 27
muveeyo [-va-yo] *verb, intransitive 2pl.* **1** clear the way; **2** become visible; **3** come out / *freq.38* rank 4440 # texts 11

navaayo [**-va-yo**] *verb, intransitive 1sg.* **1** returned; **2** came back; **3** left */ freq.13* rank 10874 # texts 8

okuvaayo [**-va-yo**] *verb, intransitive* **1** come out; **2** become visible */ freq.45* rank 3858 # texts 30

twavaayo [**-va-yo**] *verb, transitive 1pl.* **1** left; **2** came out */ freq.15* rank 9746 # texts 9

vaayo [**-va-yo**] *verb, intransitive 2sg.* **1** move away; **2** come out; **3** become visible */ freq.17* rank 8765 # texts 15

waavaayo [**-va-yo**] *verb, intransitive cl. 1* **1** come out; **2** become visible; **3** move away */ freq.13* rank 11077 # texts 10

yaavaayo [**-va-yo**] *verb, transitive cl. 1* came out */ freq.22* rank 7029 # texts 15

yavaayo [**-va-yo**] *verb, transitive cl. 1* **1** left; **2** came out; **3** became visible */ freq.18* rank 8311 # texts 13

Types in the 1.7m Lusoga corpus considered for the lemma -*v*o in cl. 14, dev. noun:

buvo [**-vo**] *noun -/14* place of origin */ freq.21* rank 7085 # texts 14

obuvo [**-vo**] *noun -/14* origin */ freq.19* rank 7858 # texts 9