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 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’eiitu 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’eiitu ogukozesebwa mu namawanika w’ennimi dha Bantu ogulaga omulimu ogw’akolebwa ku Lusoga. Oluvainhuma lw’okwandhula engeli eiitu ly’Olusoga mu Kitundu 1 n’omuteeko gw’omutindigio ogusinzilwaku okuzimba oluuka lwa namungi w’ebigambowazo mu Kitundu 2, buti eisila liize ku kulaga ngeli amakulu g’ebigambo ye gakwananganizibwa n’enkozesa. Omusingi gw’eisomo eiligibwa mu kitundu kino gw’ateebwawo Patrick Hanks. Ensongo enkulu dhe yataaku eisila dhilondeebwa okusinzilwa ku kigelo kya namungi w’enningizo entabaazi (o)ku.v.a. Olupapula luno lugelalagelania engeli ennningizo eno bwe y’alingizibwa mu Eiwanika ly’Olusoga elitaasinziilwa ku itu lya bigambo n’engeli gye yandibaela esengekebwa singa eiitu n’ebigelo by’emiwendo egilagibwamu byali bikozeseibwa. Eby’asoboka n’ebitaasoboka bigelaagelanihizibwa n’ebigendelelewa by’omusingi gw’eisomo ly’eiitu lya namawanika mu mpandiika y’amawanika.

Ebigambo ebikulu: BANTU, LUSOGA, EITU LYA NAMAWAIKA, ENNEKEENENIA Y’EBIGELO BY’EMIWENDO EBIGAMBO BA ITU LY’OULULUMI, OKUKWANAGANIA AMAKULU N’ENKOZESA, AMAKULU AGASOBOKA, KINANTABILA OMUTABAABAZI

Lexikos 28 (AFRILEX-reeks/series 28: 2018): 112-151
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, Faas 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:

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

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:

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

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 bottleneck. 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 component 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 Taljard 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-
your 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).

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

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

<table>
<thead>
<tr>
<th>e-Eiwanika</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(o)ku.v(^{a}) [(o)kůvá] kt.[L] [-viile] [viile] bl: [Lg: okuva]</td>
<td>[lemma sign, part of speech, morphological information, cognate in Luganda]</td>
</tr>
<tr>
<td>1. Okusimbuka mu kifo ekilala waayolekela ekindi. gez: Nva Mayuge.</td>
<td>1. to depart. e.g. <em>I come from Mayuge.</em></td>
</tr>
<tr>
<td>2. Okusibuka. gez: Nva Iganga.</td>
<td>2. to come from. e.g. <em>I hail from Iganga.</em></td>
</tr>
<tr>
<td>3. Okulekelela ekintu ky’obaile okola. gez: Ebyo nabiviileku.</td>
<td>3. to abandon. e.g. <em>I gave up on those things.</em></td>
</tr>
<tr>
<td>4. Okusegukula mu ngila oba mu kifo. gez: Leka nkviile ofune eidembe.</td>
<td>4. to make way. e.g. <em>Let me pave the way for you so that you get peace.</em></td>
</tr>
<tr>
<td>[combination(s) with the lemma]</td>
<td></td>
</tr>
<tr>
<td>● Okuva ku luguudo: Okwonooneka / Okuva ku mulume</td>
<td>● to go off the road: to be completely damaged and unusable / to be out of fashion</td>
</tr>
<tr>
<td>[proverbs]</td>
<td></td>
</tr>
<tr>
<td>♦ Awavva akwita n’awava akukobela</td>
<td></td>
</tr>
<tr>
<td>♦ Awava ennume waila nnume</td>
<td></td>
</tr>
<tr>
<td>♦ Awava mwino tiwaila mwino: Awava elino waila iilbu</td>
<td></td>
</tr>
<tr>
<td>♦ Awava mwino tiwaila mwino: Awava elioso waila itulu</td>
<td></td>
</tr>
<tr>
<td>♦ Akaviile mu igi tikatya ikoli</td>
<td></td>
</tr>
<tr>
<td>♦ Edhiva okulala n’embilo</td>
<td></td>
</tr>
<tr>
<td>♦ Empambo eva ku kiwalo</td>
<td></td>
</tr>
<tr>
<td>♦ Ennhonhi eva ewala temala mutonto</td>
<td></td>
</tr>
<tr>
<td>♦ Ensanafu eva ku mugendelo telwa kufuka kabasa</td>
<td></td>
</tr>
<tr>
<td>♦ Atava ku mulungi afa t’awoza</td>
<td></td>
</tr>
<tr>
<td>♦ Ka nduviile ku ntobo oti n’onuyala atuse we bafumba</td>
<td></td>
</tr>
<tr>
<td>♦ Olusubi olulala we luva ku ndhu tetoonha</td>
<td></td>
</tr>
<tr>
<td>♦ Omukazi omulungi nnimilo ya ngila buli avaayo agyeegwaniza</td>
<td></td>
</tr>
<tr>
<td>ssk:</td>
<td></td>
</tr>
<tr>
<td>● Okuva ku luguudo: Okwonooneka / Okuva ku mulume</td>
<td></td>
</tr>
<tr>
<td>(g)gl:</td>
<td></td>
</tr>
<tr>
<td>♦ The person who warns you comes from the same place as the person who will kill you</td>
<td></td>
</tr>
<tr>
<td>♦ Where a male leaves another male will take over that place</td>
<td></td>
</tr>
<tr>
<td>♦ The gap that your friend leaves is not filled by another friend: A gap takes the place of a tooth that has left</td>
<td></td>
</tr>
<tr>
<td>♦ The gap that your friend leaves is not filled by another friend: Blindness takes the place of the eye that has left</td>
<td></td>
</tr>
<tr>
<td>♦ The one that has just come from an egg does not fear an eagle</td>
<td></td>
</tr>
<tr>
<td>♦ The steps you take one after the other develop into running</td>
<td></td>
</tr>
<tr>
<td>♦ A wise lesson is learned from the cradle</td>
<td></td>
</tr>
<tr>
<td>♦ The bird that comes from far away does not finish up the edible fruit</td>
<td></td>
</tr>
<tr>
<td>♦ The safari ant that leaves the trail does not take long to turn into a traitor</td>
<td></td>
</tr>
<tr>
<td>♦ The one who does not let a beautiful one alone dies while still giving explanations</td>
<td></td>
</tr>
<tr>
<td>♦ Let me start from the very beginning like the hungry person who has arrived at the place where food is being cooked</td>
<td></td>
</tr>
<tr>
<td>♦ When one blade of grass falls off the house the house does not leak</td>
<td></td>
</tr>
<tr>
<td>Bantu Phrase</td>
<td>English Translation</td>
</tr>
<tr>
<td>--------------------</td>
<td>----------------------------------------------------------</td>
</tr>
<tr>
<td>♦ Omukwano guva mu ngabo</td>
<td>A beautiful woman is a garden along the road: whoever comes by wants it for himself</td>
</tr>
<tr>
<td>♦ Omukwano guva mu ngila gwatuka eka</td>
<td>Friendship comes from the shield (sharing)</td>
</tr>
<tr>
<td>♦ Omusaadha kikele kiva kyonka mu bwina</td>
<td>Friendship comes from the road and it is brought home</td>
</tr>
<tr>
<td>♦ Va we ndi takulwania</td>
<td>A man is a frog, which comes out of the hole by itself</td>
</tr>
<tr>
<td>♦ Va ku ntebe ya lata awulilila ku ise</td>
<td>(The one who says that) ‘Go away from where I am’ should not make you fight</td>
</tr>
<tr>
<td>♦ W’ova tosoile w’otela okwila</td>
<td>(The one who says that) ‘Go away from father’s chair’ has heard it from his father</td>
</tr>
<tr>
<td>♦ A beautiful woman is a garden along the road: whoever comes by wants it for himself</td>
<td></td>
</tr>
<tr>
<td>♦ Friendship comes from the shield (sharing)</td>
<td></td>
</tr>
<tr>
<td>♦ Friendship comes from the road and it is brought home</td>
<td></td>
</tr>
<tr>
<td>♦ A man is a frog, which comes out of the hole by itself</td>
<td></td>
</tr>
<tr>
<td>♦ (The one who says that) ‘Go away from where I am’ should not make you fight</td>
<td></td>
</tr>
<tr>
<td>♦ (The one who says that) ‘Go away from father’s chair’ has heard it from his father</td>
<td></td>
</tr>
<tr>
<td>♦ The place that you leave without picking a quarrel is the one where you always return</td>
<td></td>
</tr>
</tbody>
</table>

### (okuva) 2

#### [ lemma sign, part of speech, morphological information, cognate in Luganda ]

<table>
<thead>
<tr>
<th>Bantu Phrase</th>
<th>English Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>♦ Okutanikila mu kifo ekilala okutuuka ku kifo ekindi. gez: Ennhandha ya Nalubaale eva Idindi.</td>
<td>to start at a given point and move in the direction of another. e.g. Lake Victoria starts in Jinja.</td>
</tr>
</tbody>
</table>

### (okuvaak[A])

#### [ lemma sign, part of speech, morphological information, cognate in Luganda ]

<table>
<thead>
<tr>
<th>Bantu Phrase</th>
<th>English Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>♦ Okutandiikila mu kifo ekilala okutuuka ku kifo ekindi. gez: Ennhandha ya Nalubaale eva Idindi.</td>
<td>to start at a given point and move in the direction of another. e.g. Lake Victoria starts in Jinja.</td>
</tr>
</tbody>
</table>

### (okuvaak[A]mu) 2

#### [ lemma sign, part of speech, morphological information, cognate in Luganda ]

<table>
<thead>
<tr>
<th>Bantu Phrase</th>
<th>English Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>♦ Obutatuukiliza kye wasuubiza omuntu. gez: Tubaile tusuubiila nti agya kutuyamba aye atuviilemu.</td>
<td>to not fulfil what is expected of you. e.g. The one who promised to help us has failed us.</td>
</tr>
<tr>
<td>♦ Okulyamu oluwe. gez: Gvetwateesa naye mwene neeyatuvaamu.</td>
<td>to betray. e.g. The actual person we planned with is the one who betrayed us.</td>
</tr>
<tr>
<td>♦ Okuwa oba okumaliliza. gez: Bw’obifumba bivaamu bulungi.</td>
<td>to turn out well. e.g. When you cook them they come out very well.</td>
</tr>
<tr>
<td>♦ A beautiful woman is a garden along the road: whoever comes by wants it for himself</td>
<td></td>
</tr>
<tr>
<td>♦ Friendship comes from the shield (sharing)</td>
<td></td>
</tr>
<tr>
<td>♦ Friendship comes from the road and it is brought home</td>
<td></td>
</tr>
<tr>
<td>♦ A man is a frog, which comes out of the hole by itself</td>
<td></td>
</tr>
<tr>
<td>♦ (The one who says that) ‘Go away from where I am’ should not make you fight</td>
<td></td>
</tr>
<tr>
<td>♦ (The one who says that) ‘Go away from father’s chair’ has heard it from his father</td>
<td></td>
</tr>
<tr>
<td>♦ The place that you leave without picking a quarrel is the one where you always return</td>
<td></td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th>Bantu Phrase</th>
<th>English Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>♦ Okuviila, Okuviisa.</td>
<td>to go/turn off. e.g. The electricity has gone off.</td>
</tr>
<tr>
<td>♦ Okuviilikwaku, Okuviisaaku.</td>
<td>1. to go/get.</td>
</tr>
<tr>
<td>♦ Okupa, [okuvaak] 2 [kt.[L] -viile] [nvilee] bl: [Lg: okuva]</td>
<td>2. to let alone/put aside. e.g. I have put alcohol drinking aside.</td>
</tr>
<tr>
<td>♦ Okuvaako</td>
<td>3. to not be properly fixed. e.g. These shoes do not fit.</td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th>Bantu Phrase</th>
<th>English Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>♦ Obutatuukiliza kye wasuubiza omuntu. gez: Tubaile tusuubiila nti agya kutuyamba aye atuviilemu.</td>
<td>to not fulfil what is expected of you. e.g. The one who promised to help us has failed us.</td>
</tr>
<tr>
<td>♦ Okulyamu oluwe. gez: Gvetwateesa naye mwene neeyatuvaamu.</td>
<td>to betray. e.g. The actual person we planned with is the one who betrayed us.</td>
</tr>
<tr>
<td>♦ Okuwa oba okumaliliza. gez: Bw’obifumba bivaamu bulungi.</td>
<td>to turn out well. e.g. When you cook them they come out very well.</td>
</tr>
<tr>
<td>♦ A beautiful woman is a garden along the road: whoever comes by wants it for himself</td>
<td></td>
</tr>
<tr>
<td>♦ Friendship comes from the shield (sharing)</td>
<td></td>
</tr>
<tr>
<td>♦ Friendship comes from the road and it is brought home</td>
<td></td>
</tr>
<tr>
<td>♦ A man is a frog, which comes out of the hole by itself</td>
<td></td>
</tr>
<tr>
<td>♦ (The one who says that) ‘Go away from where I am’ should not make you fight</td>
<td></td>
</tr>
<tr>
<td>♦ (The one who says that) ‘Go away from father’s chair’ has heard it from his father</td>
<td></td>
</tr>
<tr>
<td>♦ The place that you leave without picking a quarrel is the one where you always return</td>
<td></td>
</tr>
</tbody>
</table>

---

### (okuvaak[A]mu) 2

#### [ lemma sign, part of speech, morphological information, cognate in Luganda ]

<table>
<thead>
<tr>
<th>Bantu Phrase</th>
<th>English Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>♦ Okuviila, Okuviisa.</td>
<td>1. to go/turn off. e.g. The electricity has gone off.</td>
</tr>
<tr>
<td>♦ Okuviilikwaku, Okuviisaaku.</td>
<td>2. to let alone/put aside. e.g. I have put alcohol drinking aside.</td>
</tr>
<tr>
<td>♦ Okuvaako</td>
<td>3. to not be properly fixed. e.g. These shoes do not fit.</td>
</tr>
</tbody>
</table>

---

### (okuvaak[A]mu) 2

#### [ lemma sign, part of speech, morphological information, cognate in Luganda ]

<table>
<thead>
<tr>
<th>Bantu Phrase</th>
<th>English Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>♦ Okuviila, Okuviisa.</td>
<td>1. to go/turn off. e.g. The electricity has gone off.</td>
</tr>
<tr>
<td>♦ Okuviilikwaku, Okuviisaaku.</td>
<td>2. to let alone/put aside. e.g. I have put alcohol drinking aside.</td>
</tr>
<tr>
<td>♦ Okuvaako</td>
<td>3. to not be properly fixed. e.g. These shoes do not fit.</td>
</tr>
</tbody>
</table>

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

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

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 okua (freq. 2,668), ova (freq. 389), ovaa (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 -vii- 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.)

*okuva*1 [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 bufulumu 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-funze-e
   SM₂-APPL-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.*
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.  

'...move away even from my presence; let us not kick each other from here ...'  

What came out of the examination ...'  

'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.'
okuva [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 kivilla ilala ku Bbaibbuli wano Yesu bwe yagyanga nga alonda abayigilizwa yaabaagaananga na ki? Ni profession dhaibwe.

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

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

Combinations

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

-v- + LOC + maiso = to die

Nga kitalo muna omukaile okutuva ku maiso!

As it is sad my counterpart for the old person to die!
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 \textit{kuv\_a} itale ghano ghati \textit{okutuuk\_ila} ghano.\]

\(\text{ni} \ \text{aba-li} \ \text{kuyonoon-a} \ \text{kuv-a} \ \text{itale} \ \text{gha-no} \ \text{ghati} \ \text{oku-tuuk-il-a} \ \text{gha-no} \)

\([...\] they are the ones who are spoiling \textit{from} above here like this \textit{up to} here.’

[File ID: Okukyal | O • Celebrations • Politics • 2011]

-\(v\) + \textit{paka} = from ... up to, from ... until [measuring time]

\(\text{Nkola kuv\_a} \ \text{saawa ina} \ \text{paka} \ \text{musanvu ogw’obwile}.\)

\(\text{n-kol-a} \ \text{kuv-a} \ \text{saawa ina} \ \text{paka} \ \text{musanvu o-gu-a} \)

\([...\] they work \textit{from} 10:00 \textit{until} 1:00 o’clock in the night.’

[File ID: BuwaabGr | O • Celebrations • Inspirational • 2010]

\textbf{Other word classes}

Addendum 2 indicates that, among the formatives of the verb -\(v\)-, one also finds the forms \textit{kyava}, \textit{kyebaava}, \textit{kyenva}, \textit{kyetuva} and \textit{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.

\textit{kye}-SM-v-a (connective) = that is why

\(\text{Buti \ \textit{kyenva} \ \text{tuayenze} \ \text{kufuna} \ \text{batoototo} \ \text{kuba} \ \text{boona} \ \text{baidha} \ \text{kuba} \ \text{bakyotala} \ [...]}\)

\(\text{buti} \ \text{kye-n-v-a} \ \text{ti-a-yend-ile} \ \text{kuv-fun-a} \)

\(\text{now} \ \text{NEG}\_\text{-SM\_result-FV} \ \text{NEG\_PST\_want-PERF} \ \text{15-get-FV} \)

\(\text{ba-toototo} \ \text{kuba} \ \text{ba-ona} \ \text{ba-idh-a} \ \text{kub-a} \ \text{ba-kytalana} \)

\(\text{AP\_young because PP\_INC SM\_come-FV 15-be-FV 2-half-caste} \)

\(\text{’Now \ that is (the reason) why [...]} \ \text{did not want to get young ones because even they will be half-castes [...]}\)

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

From -\(v\)-2, sense 1, the adverb \textit{kuv\_a} is derived.

\textit{kuv\_a} (adverb) = since

\(\text{Nnhweileku aye kuv\_a nkyo nkaali kulyaku}.\)
The locativised verb -vaawo

When the class 16 locative enclitic -wo is suffixed to the base verb -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).

1. to stop existing, to die (out)
   Ogenda toyenda eliyo ebiidha okuvaawo.
   o-yend-a ti-o-yend-a e-li-yo ebi-idh-a
   SM2SG-want-FV NEG1SG,SM2SG-want-FV SM23-be-ENCL23 SREL4-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-NEG8-go_away-FV-ENCL16
   tu-li ku-il-a
   SM2PL-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]

The locativised verb -vaaku

When the class 17 locative enclitic -ku is suffixed to the base verb -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. 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 ga-vile-ku [...] SM-go-off-PERF-ENCL
Teh! The electricity has gone off [...] [File ID: PFExtaud | O • Interviews • Language • 2012]

2. to abandon
Oba ti na kindi okutuviilaku ilala [...] CP
oba ti-na-ki-ndi oku-tu-v-il-a-ku
or NEG-MODF-PP-EXC 15-OM-pl-abandon-APPL-FV-ENCL 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 okutuluugunhizibwa mnankani, buvunaanhizibwa buife.
eki-li-Ø-v-il-a-ku ba-ana
SREL-PROG-15-cause-APPL-FV-ENCL 2-child
oku-tuluugunh-is-ibw-a mnankani bu-vunaanhizibwa bu-a-ife
15-abuse-CAUS-PASS-FV UNS 14-responsibility CP-of-PRON-pl
What is causing the children to be abused, (is) something (that is) our responsibility.’ [File ID: PI101216 | O • Radio talk shows • Health • 2010]

4. to let aside, to give up
Maana wange, kulwaki tobivaaku?
u-mu ana a-a-nge ku-lu-a-ki
1-child CP-of-POSS-SSG PP-CP-of-INTER
ti-o-bi-v-a-ku NEG-APL-sm-SSG-OM-give_up-FV-ENCL
'My child why don't you give them up?'

5. to lose
Tusaasilaku abantu abaaviilwaku abantu baibwe modulo oti n’e yo.

6. to stop
ARVs waadhivaaku?

7. to not disturb, to leave alone
Leave me alone you young boy; but really why do you stand against me in that way?

8. to finish
After the mothers have finished (giving their speeches), we shall be with the words emanating from the fathers.
9. **to come a (little) bit**

Nga bwe tu-a-ba-tegeez-is-a oku-v-a eila nti munange tugya kuba n’abakungu okuvaaku mu 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.

\[-vaaku omusolo = to remit tax\]

Amakolelo a-ga-vaaku omusolo.

AUG-6-do-RPT-DEV SREL₄-let_aside-FV-ENCL₁₇ 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 *-v₁*, numerous 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. 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 ebi-balile bili kwogelwa nti akalulu akaviilemu [...]

AUG-6-do-RPT-DEV SREL₄-let_aside-FV-ENCL₁₇ AUG-3-tax

[File ID: Judicial | O • Radio talk shows • Sensitization • 2010]
2. to come out, to flow out, to exit

Owundi ku baisilukale yaamufumita eifumo mu lubavu ela mangu ago muavaamu omusaayi n’amaadih.  
o-wu-ndi-ku ba-isilukale a-a-mu-fumit-a e-i-fumo  
AUG-PP$_1$-EXC-ENCL$_{17}$ 2-soldier SM$_1$-PST-OM$_1$-pierce-FV AUG-5-spear  
mu lu-bavu ela mangu a-ga-o mu-a-v-a-mu  
LOC$_{18}$ 11-rib CON quickly AUG-PP$_2$-DEM$_8$ SM$_{15}$-PST-come-FV-ENCL$_{16}$  
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 flew out.’

[File ID: Missa1 | W • Biblical documents • Religion • 2012]

3. to grow well, to turn out well

Bw’ozada n’abaawo ekivaamu agasa bwa iwanga.  
bwe o-zaa-l-a ni a-b-a-wo  
if SM$_{12}$-give_birth-FV CON SM$_1$-be-FV-ENCL$_{16}$  
ek-i-v-a-mu a-gas-a bwa i-wanga  
SREL$_2$-turn_out_well-FV-ENCL$_{16}$ SM$_1$-benefit-FV just 5-country  
‘If you give birth and s/he turns out well, s/he just benefits the nation.’

[File ID: Bwozaal | O • Songs - Traditional • Inspirational • 2000s]

4. to yield, to generate

Ekigihulo kya Kyabazinga kyatundibwa aghalala ni Kodh’eyo kyavaamu emitwalo kumpi ikumi nga empiiya dhiho dha kugheeleza Busoga mu bitundu ebili n’eb’etaago.  
e-ki-ighul-o ki-a kyabazinga ki-a-tund-ibw-a aghalala  
AUG-7-food-DEV CP$_7$-of 1a.King SM$_2$-PST-sell-PASS-FV together  
ni kodh’eyo ki-a-v-a-mu e-mi-twalo kumpi  
CON 9.Kodh’eyo SM$_{18}$-PST-yield-FV-ENCL$_{16}$ AUG-4-ten_thousand near  
ikiumi nga e-n-piiya dhi-no dhi-a ku-gheel-is-a  
ten CON AUG-10-money PP$_{10}$-DEM$_A$ CP$_{10}$-of 15-serve-CAUS-FV  
bi-soga mu bi-tundu ebli ni e-bi-etaag-o  
14-soga LOC$_{18}$ 8-part SREL$_{6}$-be CON AUG-8-need-DEV  
‘(Tickets to attend) the ‘dinner’ (in honour) of the King were sold together with Kodh’eyo and this yielded roughly one hundred thousand Shillings which serves Busoga in the parts which have needs.’

[File ID: Kodh’eyo | W • Journalism • Networking • 1997–1998]
Corpus-driven Bantu Lexicography: Mapping Meaning onto Use in Lusoga

5. to not gain

Muzeeyi iwe oidha kwiilamu awo.

mu-zeeyi iwe o-idh-a ku-v-il-a-mu
1-old PRON2SG SM2SG-will-FV 15-not_gain-APPL-FV-ENCL18
a-wa-o
AUG-PP1o-DEM8

'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' auaamu enda. Onaagya waalima?

mu-kazi a-o lwe a-v-a-mu en-da
1-wife PP1-DEM8 OREL11 SM1-come_out-FV-ENCL18 10-stomach
o-naa-gi-a o-a-lim-a
SM2SG-FUT1-go-FV SM2SG-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 bwelayagala; Wakayima, yaasalawo okuya okwekweka mu ndhu.

eki-a-v-a-mu wa-mpala a-a-koogh-a oku-tambul-a
SREL-PST-come_out-FV-ENCL18 16-lion SM1-PST-tire-FV 15-walk-FV
nga bwe a-a-gal-a wa-kayima a-a-sal-a-wo
ADV ADV SM1-PROG-search-FV 16-monkey SM1-PST-decide-FV-ENCL16
oku-gi-a oku-e-kwek-a mu n-dhu
15-go-FV 15-RFL-hide-FV LOC18 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. to come out
2. to give way

1. to come out

2. to give way
Eee! Oooh! vaayo baiddhaakuniina. eee oooh v-a-yо ba-idh-a Ø-ku-niin-a INTERJ INTERJ give_way-FV-ENCL23 SM2-will-FV 15-OM25C-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

Inhonholo obuvo n'obwiko bve ensonga yaali kwogelaku.
inohnhol-o-a o-bu-v-o ni o-bu-ik-o bu-a explain-FV AUG-14-start-DEV and AUG-14-end-DEV CP14-of e-n-songa ye a-li ku-yogel-a-ku AUG-9-issue OREL9 SM1-be 15-speak-FV-ENCL17 '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 [ ].)

<table>
<thead>
<tr>
<th>Lemma signs</th>
<th>Meaning potentials</th>
<th>Combinations, + meaning potentials following ‘=’</th>
<th>Lemma signs for other word classes, + meaning potentials following ‘=’</th>
</tr>
</thead>
<tbody>
<tr>
<td>okuvanz — [move senses]</td>
<td>1. to leave, to depart, to go away [1]</td>
<td>-v- + LOC + maiso = to die</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2. to hail (from) [2]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3. to abandon [3]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4. to make way, to move away [4]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5. to result, to come out</td>
<td></td>
<td>kye-SM-va (connective) = that is why</td>
</tr>
<tr>
<td></td>
<td>6. to spend (time)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>okuvanz — [projection and direction senses]</td>
<td>1. to start (and continue onwards) [1]</td>
<td>-v- + -luuk- = from up to [measuring space]</td>
<td>kuru (adverb) = since</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-v- + paka = from ... up to, from ... until [measuring time]</td>
<td></td>
</tr>
</tbody>
</table>
2. to be the source
(of), to emanate
(from)

okuvaawo < okuva⁴
1. to stop existing, to
die (out)
2. to leave, to depart,
to go away

okuvaaku < okuva⁴
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 < okuva⁴
1. to abandon though
   it is expected [1]
2. to come out, to
   flow out, to exit [5]
   -vaamu enda = to
   miscarry
   ekvama = 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 < okuva⁴
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 locativised 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. *okwa 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. 1 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 omuwoosi wawa 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. 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 nevertheless 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

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>#</td>
<td>noun class number</td>
</tr>
<tr>
<td>ADV</td>
<td>adverb</td>
</tr>
<tr>
<td>AFP</td>
<td>affiliation prefix</td>
</tr>
<tr>
<td>APERF</td>
<td>after-perfect</td>
</tr>
<tr>
<td>APPL</td>
<td>applicative</td>
</tr>
<tr>
<td>AP</td>
<td>adjectival prefix (of cl. x)</td>
</tr>
<tr>
<td>AUG</td>
<td>augment</td>
</tr>
<tr>
<td>CAUS</td>
<td>causative</td>
</tr>
<tr>
<td>cl.</td>
<td>class</td>
</tr>
<tr>
<td>CON</td>
<td>connective</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>COP</td>
<td>copulative</td>
</tr>
<tr>
<td>CP_x</td>
<td>connective prefix (of cl. x)</td>
</tr>
<tr>
<td>DEM_A,B,C</td>
<td>demonstrative (of position A, B, C)</td>
</tr>
<tr>
<td>DEV</td>
<td>deverbative</td>
</tr>
<tr>
<td>ENCL_x</td>
<td>locative enclitic (x = cl. 16, 17, 18, 23)</td>
</tr>
<tr>
<td>EXC</td>
<td>exclusive pronoun</td>
</tr>
<tr>
<td>ext.</td>
<td>extension</td>
</tr>
<tr>
<td>FUT_1</td>
<td>future tense -nää-</td>
</tr>
<tr>
<td>FUT_2</td>
<td>future tense -li-</td>
</tr>
<tr>
<td>FV</td>
<td>final vowel</td>
</tr>
<tr>
<td>HAB</td>
<td>habitual</td>
</tr>
<tr>
<td>INC</td>
<td>inclusive pronoun</td>
</tr>
<tr>
<td>INTENS</td>
<td>intensifier</td>
</tr>
<tr>
<td>INTER</td>
<td>interrogative</td>
</tr>
<tr>
<td>INTERJ</td>
<td>interjection</td>
</tr>
<tr>
<td>loc.</td>
<td>locativised</td>
</tr>
<tr>
<td>LOC_x</td>
<td>locative (x = cl. 16, 17, 18, 23)</td>
</tr>
<tr>
<td>MOD</td>
<td>modality</td>
</tr>
<tr>
<td>MODF</td>
<td>modifier</td>
</tr>
<tr>
<td>NEG_A,B</td>
<td>negative (of type A, B)</td>
</tr>
<tr>
<td>OM_x</td>
<td>object marker (of cl. or ps. x)</td>
</tr>
<tr>
<td>OREL_x</td>
<td>object relative (of cl. x)</td>
</tr>
<tr>
<td>PASS</td>
<td>passive</td>
</tr>
<tr>
<td>PERF</td>
<td>perfect</td>
</tr>
<tr>
<td>PL</td>
<td>plural</td>
</tr>
<tr>
<td>POSS_x</td>
<td>possessive (of ps. x)</td>
</tr>
<tr>
<td>PP_x</td>
<td>pronominal pronoun (of cl. x)</td>
</tr>
<tr>
<td>PROG</td>
<td>progressive</td>
</tr>
<tr>
<td>PRON_x</td>
<td>pronoun (of ps. x)</td>
</tr>
<tr>
<td>ps.</td>
<td>person</td>
</tr>
<tr>
<td>PST</td>
<td>past tense</td>
</tr>
<tr>
<td>REC</td>
<td>reciprocal</td>
</tr>
<tr>
<td>RFL</td>
<td>reflexive</td>
</tr>
<tr>
<td>RPT</td>
<td>repetitive</td>
</tr>
<tr>
<td>SG</td>
<td>singular</td>
</tr>
<tr>
<td>SM_x</td>
<td>subject marker (of cl. or ps. x)</td>
</tr>
<tr>
<td>SREL_x</td>
<td>subject relative (of cl. x)</td>
</tr>
<tr>
<td>SUBJ</td>
<td>subjunctive</td>
</tr>
<tr>
<td>UNS</td>
<td>unspecified noun</td>
</tr>
</tbody>
</table>

**Acknowledgements**

The research for this article was funded by the Special Research Fund of Ghent University. Thanks are due to the two anonymous referees.

**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 -caawo also has a variant, namely -caagho, but its frequency is too low to have made it into the lemmatised frequency list.
8. ARVs = antiretrovirals (i.e., drugs to treat HIV)
10. For the value of corpus examples over ‘invented’ (but more didactic) ones see Fox (1987).

References


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

(o)ku.va

(o)ku.va.1 [ (c) kuva ] k.t. [L] [ -vile ]

(gg): Awava akwana n’awava akukobela; Awava ennuye wala mume. Awava mwino twiwa mwino: awava elino wala itibwa; Awava elino wala itibwa. Akivilele mu izi tukusa kikolo; Edhiwa okulala n’embolo. Emumko eva ku kivuli.

Emanafa eva ku mugendele tuleka kufika kabasa, Atava ku mukundi afa t’awoza. Ka ndurule ku ntondo ot’w’umuya atumwe bajambwa. Okusuki okulala we kwa ku ndu tetoza; Ommake omulungi amulongo ya ngala bila avavya agyege aama. Omumwoyo guwa mu ngala; Omumwoyo guwa mu ngala guwa muganda. Eka. Omumwoyo kikele kiva skyanka mu bwaza. Va ve ada takwabana; Va ku nizee ya lata awilisha ku ise; W’eva twiwa w’otela okula.

bgg: Okuviula, Okuviula.

(o)ku.va.2 [ (c) kuva ] k.t. [L] [ -vile ]

(o)ku.va.bila.1 [ (c) kuva.bila ] k.t. [L] [ -vile ]
[vile] [vile] bl: [Lg: okuva] 1) Okubisa onumbwe, emakono, amagulu oba embile mu k’u kigomye oba engato. ge: Aamuntu bwinisa engoye balaka bahita wadu.

bgg: Okuva.bila, Okuva.bila, Okuva.bila, Okuva.bila, Okuva.bila.

(o)ku.va.2 [ (c) kuva ] k.t. [L] [ -vile ]

bgg: Okuva.bila, Okuva.bila, Okuva.bila, Okuva.bila, Okuva.bila.
Addendum 2:  Top orthographic corpus types in the 1.7m Lusoga corpus underlying the verbal lemmas -vo-, -voavo (cl. 16), -voaku (cl. 17), -voamu (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 -vo-, verb:

abaava [-va] verb, transitive cl. 2 1 came from; 2 diverted / freq.70 rank 2606 # texts 22;
abava [-va] verb, transitive cl. 2 came from / freq.81 rank 2289 # texts 32;
abaville [-va] verb, transitive cl. 2 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;
awa [-va] verb, transitive cl. 1 from / freq.39 rank 509 # texts 107;
ave [-va] verb, intransitive cl. 1 1 comes from; 2 leaves / freq.37 rank 4478 # texts 17;
aville [-va] verb, transitive cl. 1 came from; 2 diverted; 3 left / freq.35 rank 4655 # texts 25;
aviire [-va] verb, transitive cl. 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;
bava [-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;
baville [-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);
byava [-va] verb, transitive cl. 8 1 from; 2 come from / freq.63 rank 2851 # texts 35;
biville [-va] verb, transitive cl. 8 1 came from; 2 diverted / freq.104 rank 1787 # texts 7;
byava [-va] verb, transitive cl. 14 from come from / freq.63 rank 3296 # texts 35;
byava [-va] verb, transitive cl. 6 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;
ebava [-va] verb, transitive cl. 8 come from / freq.106 rank 1753 # texts 35;
ebava [-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;
ekiva [-va] verb, transitive cl. 7 1 came from; 2 left / freq.18 rank 8060 # texts 7;
evava [-va] verb, transitive cl. 9 from / freq.118 rank 1602 # texts 44;
eyava [-va] verb, transitive cl. 7 came from / freq.46 rank 3748 # texts 21;
gava [-va] verb, transitive cl. 6 came from / freq.33 rank 4920 # texts 25;
guava [-va] verb, transitive cl. 3 came from / freq.27 rank 5809 # texts 16;
guava [-va] verb, transitive cl. 7 1 came from; 2 left / freq.13 rank 10656 # texts 11;
kava [-va] verb, transitive cl. 12 1 came from; 2 departs / freq.13 rank 10693 # texts 9;
kikvire [-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;
kiva [-va] verb, transitive 1 leave; 2 let alone / freq.267 rank 749 # texts 96;
kiva [-va] verb, transitive 1 came from; 2 resulted from / freq.20 rank 7437 # texts 4;
kalyava [-va] verb, transitive cl. 7 1 came from; 2 diverted / freq.51 rank 3434 # texts 24 (Notes: see also kye & ava > ky’ava);
kiveba [-va] / freq.13 rank 7772 # texts 5 (Notes: see kye & baava);
kvenva [-va] / freq.36 rank 4597 # texts 17 (Notes: see kye & nva);
kuyeva [-va] / freq.13 rank 10766 # texts 10 (Notes: see kye & tuva);
kveyava [-va] verb, transitive 7 + 7 that is why / freq.143 rank 1329 # texts 16 (Notes: see kye & yava);
 livava [-va] verb, transitive cl. 11 comes from / freq.32 rank 5052 # texts 15;
liva [-va] verb, transitive cl. 11 1 comes from; 2 results from / freq.20 rank 7443 # texts 12;
liva [-va] verb, transitive cl. 11 came from / freq.17 rank 8589 # texts 6;
muava [-va] verb, transitive 2pl. come from / freq.30 rank 5351 # texts 15;
muava [-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;
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 1

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 depart; 2 results into; 3 releases / freq.18 rank 7990 # texts 15

**bituviiremu** [va-mu] verb, intransitive cl. 8 + lpl. 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 come 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, transitive cl. 1 1 comes out; 2 surfaces / freq.35 rank 4654 # texts 25

**aveeayo** [va-yo] verb, intransitive cl. 1 1 move away; 2 become visible / freq.14 rank 9851 # texts 13

**baveeyo** [va-yo] verb, transitive 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 1sp. 1 returned; 2 came back; 3 left / freq. 13 rank 10874 # texts 8
okuvaayo [-va-yo] verb, intransitive 1 came 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 come out; 2 become visible; 3 move away / freq. 13 rank 11077 # texts 10
yaavaayo [-va-yo] verb, transitive cl. 1 left; 2 came out / freq. 22 rank 7029 # texts 15
yavaayo [-va-yo] verb, transitive cl. 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 -vo 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