

Diachronic semantics of the modal verb *-sóból-* in Luganda

A corpus-driven approach*

Deo Kawalya^{i, ii}, Koen Bostoen^{i, iii}
and Gilles-Maurice de Schryver^{i, iv}

ⁱ Ghent University / ⁱⁱ Makerere University /

ⁱⁱⁱ Université libre de Bruxelles / ^{iv} University of Pretoria

In this article the semantic evolution of the modal verb *-sóból-* in Luganda is examined through the analysis of a diachronic corpus spanning the last 120 years. It is shown that the usage frequency of this verb increased exponentially between the early 20th century and the present, and that its use as a lexical verb has dramatically decreased during the same time span. The corpus data furthermore substantiate that its modal uses have been extended from the exclusive expression of dynamic possibility until the 1960s to deontic possibility. Methodologically, this is the first full-scale article within the young field of Bantu corpus linguistics in which a corpus-driven study is performed on a diachronic corpus. The approach is not only shown to be feasible, but the encouraging results will hopefully entice colleagues working on the Bantu languages to undertake the building and analysis of such diachronic corpora for linguistic purposes.

Keywords: modality, possibility, grammaticalisation, diachronic corpus analysis, Bantu

* The research presented in this article was supported by the Belgian Development Cooperation, VLIR-UOS (<http://vliruos.be/>), through an exceptional Short Research Stay, as well as the Special Research Fund of Ghent University (<http://ugent.be/>), through a doctoral scholarship, both awarded to the first author, with the other authors as supervisors.

1. Bantu corpus linguistics

Bantu corpus linguistics is a young field, the beginnings of which can be traced back to about 15 years ago. As it stands, all the scientific contributions to date can still be enumerated. In 1999, De Schryver published a booklet on the phonetics of Cilubà in which he made use of a Cilubà corpus of 300,000 words to not only obtain a phonetic description but also to quantify the distribution of the different phone categories of the language under study. Compared to earlier manual research, De Schryver (1999) was able to make a maximum number of claims, based on a minimum number of words, about the most frequent section of a language's lexicon. In 2002, De Schryver & Gauton presented the first corpus-based diachronic analysis of a linguistic aspect of a Bantu language. Using a 5-million-word Zulu corpus they studied the distribution of the class 17 locative prefix *ku-* and its variants *kwi-* and *ko-* in the locativisation of nouns in Zulu, and were able to reach conclusions regarding the possible changes in use that these prefixes underwent through time (between the 1930s and 1990s), and also showed how these prefixes relate to each other in sub-corpora of different genres/topics. Two years later, Gauton et al. (2004) used the same corpus to examine both the intrinsic and contextual semantic import of the nominal suffix *-kazi* in Zulu. Methodologically, the field had already moved from mere distributional research, to include diachronic aspects and deep semantics. The next two years saw the publication of three more studies. In 2005, Toscano & Sewangi consulted a Swahili corpus of half a million words to discover the actual patterns that govern the use of the class 16, 17 and 18 *amba-* locative relatives in Swahili. For some reason locatives kept attracting the embryonic Bantu corpus linguistics community, as in 2006 they were again at the centre of a study: using a 5.8-million-word Northern Sotho corpus, De Schryver & Taljard (2006) gave a minute description of the semantic contents and complement requirements of the higher-order locative *n*-grams in this language, i.e. combinations of two and three locative particles (e.g. *ka kua* or *ka mo go*), a linguistic description supported by well over a hundred examples drawn from the corpus. Using the same corpus, and looking at a time span of half a century, Taljard (2006) went on to present a semantic study illustrating the historical relationship between adjectives and enumeratives in Northern Sotho.

For this entire period, thus from 1999 until 2006, we tend to agree with Taljard (2006: 170) who wrote: "the authors have succeeded in maintaining a balance between reliance on the corpus for providing them with observable data, introspection and consultation with mother-tongue speakers as research methodologies". Using Tognini-Bonelli's (2001) terminology, one would refer to these early efforts as corpus-*based* studies.

As if to prepare for the next phase, i.e. corpus-*driven* studies, or thus studies in which the corpus itself is considered to be the sole source of hypotheses about language, the subsequent study in corpus linguistics for a Bantu language only appeared following a hiatus of four years. In 2010, and literally starting from scratch, De Schryver & Nabirye presented a corpus-driven analysis of the morphology, morphophonology and semantic import of the noun in Lusoga, a largely undocumented Bantu language, for which they first built a corpus of about 870,000 words. Most recently, the second half of the article by Bostoen et al. (2012) presented a corpus-driven synchronic study of the expression of possibility in Kirundi. The latter two studies are early results of PhD research projects, by Nabirye and Mberamihigo respectively, in which the corpora literally drive the linguistic descriptions. In the same vein, the present article is the first outcome of on-going PhD research, here by Kawalya, which is driven by corpus facts.

Some colleagues may object to the overview just presented, in that “surely there should be and there is more”. It is true that some Bantu studies have been published over the past few years in which corpus claims were made, but we feel that some kind of structured corpus must at least have been built or queried, that the authors must as a minimum be able to properly differentiate between tokens, types and lemmatised forms, and that a corpus cannot be used as a fishpond, in which to throw back what does not fit the hypothesis or theory at hand (cf. Hanks 2013: 7). Such so-called Bantu “corpus” studies have thus not been listed here.¹

Still other colleagues might argue that descriptive linguists often have quite a number of fully glossed texts in their possession, easily running into thousands of sentences, which are also “a kind of corpus”. The use of such corpora may indeed result in rich studies, such as the description of the ways in which modality is expressed in Shangaci by Devos (2008). However, a corpus for us consists of naturally produced (or translated), or naturally observed material – i.e. many thousands of words of real, running text –, not staged communicative events nor elicitations (cf. Himmelmann 1998, 2006).

In this overview, other uses of Bantu corpora in sister fields have also not been mentioned. A very productive discipline where frequent use is made of Bantu corpora is for example computational linguistics. An entire special issue of the journal *Language Resources and Evaluation* was recently devoted to the use of African language corpora in this field (De Pauw et al. 2011), to which the interested reader is referred. Other productive sister fields for the use of Bantu corpora

1. For example, Zawada & Ngcobo (2008) and Ngcobo (2010) claim to be corpus-based studies (here of the noun in Zulu), but these efforts contain important methodological flaws (amongst others by not lemmatizing the material).

include education – see for instance Prinsloo & De Schryver (2001:125–129) or Taljard (2012), both for Northern Sotho, and lexicography – see for instance De Schryver (2010), for Zulu.

2. On Luganda and modality

Luganda is the most widely spoken native language in Uganda, with well over five million first and second-language speakers (Lewis et al. 2013). It is a Great Lakes Bantu language (Schoenbrun 1994). Although Luganda has attracted the most scholarship of all Ugandan native languages, there is no study dedicated to the expression of modality. Even the most authoritative Luganda grammar just touches upon the closely related topic of “verb auxiliaries and particles” and hardly delves into the subject of modality (Ashton et al. 1954). Despite the importance of modal verbs in Bantu as exhibited in grammars, such as Bentley (1887) for Kikongo or Nash (1992) for Ruwund, and in research articles, such as Fourie (1989) for Zulu, Louwrens (1990) for Northern Sotho, Fourie (1991) for Oshindonga, Devos (2008) for Shangaci, or Bostoen et al. (2012) for Kirundi, the area remains generally understudied in Bantu languages. We are therefore conducting systematic research on modality in Luganda, the mother tongue of the first author.

In contrast to the already mentioned Devos (2008), who uses elicitation techniques to build a Shangaci corpus, and Bostoen et al. (2012), who only use a Kirundi corpus for the second part of their study, in the present article we wish to study (aspects of) modality in Luganda using corpus material only. It needs to be pointed out that such a corpus-driven semantic study requires (near-)native knowledge of the language. For languages other than those from the Bantu language family, studies of modality that are driven by corpus data only, have of course already become quite common.

Possibility is traditionally seen as one of the core components of the semantic domain of modality, along with notions such as probability, obligation, and necessity. As part of the more encompassing study of modality in Luganda, we present in this paper the results of a diachronic semantic corpus-driven analysis of the modal verb *-sóból-*. The study of this modal marker is important for at least two reasons. First of all, it is the most frequent marker of possibility in Luganda, especially in recent times. Secondly, this verb is cognate to the Kirundi possibility marker *-shóbor-*, which was analysed by Bostoen et al. (2012). Kirundi is also part of Great Lakes Bantu (Schoenbrun 1994). These two verbs, which are historically related, underwent divergent semantic changes much like the triplet *may – mögen – mogen* in English, German and Dutch respectively. Their current-day semantic range only overlaps to a certain extent.

This article is organised as follows. In Section 3 we give a general introduction to the semantic domain of modality, and its core modal concepts of possibility and necessity, as viewed by Van der Auwera & Plungian (1998) and Nuyts (2006). In Section 4 we present a detailed description of the Luganda corpus used and also say a few words about the methodology applied in sourcing the data for the present study. In Section 5 we describe the expression of possibility in Luganda with emphasis on *-sóból-* as the most frequent marker of possibility. Specifically, we shall illustrate the different uses of *-sóból-* as found in our corpus. In Section 6 we undertake a diachronic corpus analysis of the expression of possibility in Luganda, sampling three time periods, i.e. the 1900s, 1960s and 2010s. A general discussion of our findings as well as conclusions can be found in Section 7.

3. The semantic domain of modality and its core modal concepts of possibility and necessity

Modality is generally considered as a semantic subdivision within the wider TAM domain, narrowly associated with categories such as tense and aspect (Nuyts 2006: 1). Modality has not always been an easy concept to define. In its broad sense, it is used “to refer to any kind of speaker modification of a state of affairs” (Nuyts 2006: 1). Palmer (2001: 1) describes modality as being “concerned with the status of the proposition that describes the event”. It refers to the speaker’s attitude towards the action or state expressed by the main verb. Narrog (2010: 392) rather defines modality “in terms of a lack of factivity, or, from a different perspective, as the relativization of the validity of a proposition with respect to a certain background”. Although there is seemingly no consensus on the definition of the semantic domain of modality, the fact that possibility and necessity are core modal concepts is rather undisputed. This is the reason why some scholars, such as Van der Auwera & Plungian (1998), restrict the definition of modality to the expression of these core concepts.

Van der Auwera & Plungian (1998) distinguish between four modality types: (i) ‘participant-internal’; (ii) ‘participant-external’; (iii) ‘deontic’; and (iv) ‘epistemic’. They consider the first three types to be a subset of ‘non-epistemic’ and consider deontic modality to be a particular sub-type of participant-external modality. In this case they deviate from the more traditional approaches, such as the one adopted by Nuyts (2006), who distinguishes between three modality types: (i) ‘dynamic’, further subdivided into (a) ‘participant-inherent’, (b) ‘participant-imposed’, and (c) ‘situational’, thus including the first two types of Van der Auwera & Plungian (1998); (ii) ‘deontic’; and (iii) ‘epistemic’.

Dynamic modality can be characterized as capacities, abilities or potentials on the side of possibility, and needs or necessities on the side of necessity ascribed to the subject-participant of the clause, or more precisely to the controlling participant in the state of affairs, usually the agent, in order to not exclude passives (Nuyts 2006: 3). These properties can be fully inherent to the first-argument participant, as in Example (1), in which case we speak of ‘participant-inherent dynamic modality’ following Nuyts (2006) or ‘participant-internal modality’ in terms of Van der Auwera & Plungian (1998).

- (1) a. Mukasa **can** carry a sack of ninety-five kilograms.
- b. Kasozi **needs** to eat something every two hours.

If these capacities/abilities/potentials or needs/necessities are determined by external factors, which are either explicit or implicit and may be partly beyond the control of the first-argument participant, as in Example (2), we speak of ‘participant-imposed dynamic modality’ following Nuyts (2006) or ‘participant-external modality’ in terms of Van der Auwera & Plungian (1998).

- (2) a. Lydia will hire a bus, so that we **can** go to Masaka.
- b. To go to Makerere University at this hour, you **have to** take a taxi.

It is to cover cases such as those in Example (3), which go beyond capacities/abilities/potentials or needs/necessities of any participant in the state of affairs and which rather characterize a potential or a necessity/inevitability inherent in the situation described in the clause as a whole, that Nuyts (2006: 4) proposes a third dynamic sub-type, i.e. ‘situational dynamic modality’. This modal sub-type typically appears in expressions in which there simply is no participant, but also in cases with inanimate first-argument participants, and even with animate (including human) first-argument participants, in which the first-argument participant is left implicit.

- (3) a. In Uganda, it **can** be hot for six consecutive months.
- b. Pressure **must** accumulate behind the articulators for a plosive to be produced.

Although deontic modality has traditionally been defined in terms of permission and obligation (Palmer 2001: 89), Nuyts (2006: 4) prefers a more general definition, i.e. as an indication of the degree of moral desirability of the state of affairs expressed in the utterance, as in Example (4).

- (4) a. Everyone is here. You **can** start your talk now.
- b. Students **must** be silent while the teacher talks.

To Van der Auwera & Plungian (1998: 81) deontic modality is a subdomain or a special case of participant-external modality, since it “identifies the enabling or compelling circumstances external to the participant as some person(s), often the speaker, and/or as some social or ethical norm(s)”. For Nuyts et al. (2010: 17), however, this view does not take into account that deontic modality is an attitudinal category in that it indicates “the degree to which the ‘assessor’ (typically, but not necessarily, the speaker [...]) can commit him/herself to the SoA [i.e. state of affairs] in terms of certain principles”. Dynamic modality, including participant-external possibility/necessity, misses this attitudinal dimension. It simply describes possibilities or necessities inherent in agents or situations without signalling an assessment of speaker commitment to the state of affairs (Nuyts et al. 2010: 17).

The last type of modality is epistemic modality. Van der Auwera & Plungian (1998: 81) define it as “a judgment of the speaker: a proposition is judged to be uncertain or probable relative to some judgment(s)”. For Nuyts (2006: 6), it “concerns an indication of the estimation [...] of the chances that the state of affairs expressed in the clause applies in the world”. In other words, it expresses the degree of probability of the state of affairs. Uncertainty is epistemic possibility – as in Example (5a) – while epistemic necessity pertains to an event which is relatively certain by virtue of some judgment – as in Example (5b).

- (5) a. The car is not there. She **may** have gone.
- b. The bus is already here. Kajubi **must** have arrived.

Nuyts (2005, 2006) advances the ‘attitudinal’ character of both deontic and epistemic modality as the principal reason not to lump together deontic and dynamic modality, as Van der Auwera & Plungian (1998) do, and not to oppose deontic and dynamic modality together to epistemic modality, as is common in the AngloAmerican tradition of ‘root modality’ vs. epistemic modality. For Nuyts (2005, 2006) moral imposition by a subject outside the state of affairs is the dividing criterion, while Van der Auwera & Plungian (1998) take ascription to the first-argument participant to draw the line (Nuyts 2006: 7). We believe that none of these categorizations is superior *in se*. However, we do adopt here the classification proposed by Nuyts (2006), since the attitudinal dimension of deontic possibility happens to be a semantic threshold that the Luganda verb *-sóból-* only crossed recently.

4. The Luganda corpus and methodology used

Although there are over 500 Bantu languages, sizeable text corpora have only been built for about two dozen of them (De Schryver & Prinsloo 2000). Except for two POS-tagged corpora for Swahili, and three smaller POS-tagged corpora for Northern Sotho, Zulu and Cilubà, all other Bantu corpora are raw corpora, not annotated for parts of speech or any other features (De Pauw et al. 2012). Most Bantu languages are predominantly oral languages which are seldom used in written form. Corpus-building from oral data is very time consuming, so it is not surprising that sizes of Bantu corpora typically go from a few hundred thousand to a few million words only. When one sets out to build a new corpus for a Bantu language not yet covered, one typically approaches the task opportunistically, in that (i) one includes what is easily available first, (ii) taking and using complete texts wherever possible rather than equal-size samples from a much larger collection. The rationale for this approach is simply that (i) Bantu corpus linguists rarely have any funding to help in corpus building, and (ii) even if all existing texts for a certain Bantu language were included, the total size would still not go much beyond ten million words. Knowing what is in the corpus is thus more important than attempts at balance and representativeness.

In today's day and age, the obvious starting point for many under-resourced languages is the Internet. As such, the Internet is used *for* corpus building rather than *as* a corpus, a distinction first made by De Schryver (2002), who showed the potential of using the Internet for corpus building in a dozen African languages. Although it is possible to use the Internet *as* a corpus for a few Bantu languages (typically the bigger ones, e.g. Swahili or Zulu), for most other Bantu languages one will use the Internet as a source for corpus building rather, with the texts stored locally. A website which gives an indication of the likely upper limit of what's available online is Kevin Scannell's (online, see the References for the URL). The Status Page showed that his crawling software had collected about 2.25 million words for Luganda in November 2011. In contrast to Scannell's automated process, we trawled the Internet manually, selecting texts in an appropriately-written Luganda only, and collected about a million words in this way. The large difference in size between Scannell's automated collection and our manual one can be explained by the fact that we removed all sorts of (re)duplication. Scannell indeed confirmed (personal communication, e-mail of 15 November 2011) that his Luganda material contains multiple copies of the Bible. In our corpus, we kept a single copy of each of the 66 books of the Bible, good for 0.6 million words. The remaining 40% of Internet material in our corpus was selected from two online newspapers and a radio station (427 files in all), as well as from seven other sites (12 files in all).

We also obtained the electronic files of ten mostly self-published booklets, totalling about 125,000 words. Further up the difficulty cline, we set out to scan eight books, focusing on the oldest currently available printed material (from 1891 to 1926). This portion added about 250,000 words to our corpus. We also transcribed 12 songs and a radio play, adding a modest 9000 words. Finally, 27 English texts were translated into Luganda, good for another 105,000 words. Summing all of this results in a corpus of 1.5 million words (tokens), about 146,000 of which are distinct orthographic words (types). The corpus was not POS-tagged. See Table 1 for the details.

Table 1. Source distribution in the Luganda corpus

| Source | Files | Tokens | % | Types | Std TTR | Std TTR std dev |
|----------------|-------|-----------|--------|---------|---------|-----------------|
| Internet | 505 | 988,126 | 67.07 | 101,010 | 44.26 | 57.08 |
| e-Transfer | 10 | 123,490 | 8.38 | 26,890 | 56.81 | 42.97 |
| Scans + OCR | 8 | 247,721 | 16.82 | 39,286 | 50.09 | 50.27 |
| Transcriptions | 13 | 9,125 | 0.62 | 3,313 | 56.12 | 35.75 |
| Translations | 27 | 104,705 | 7.11 | 13,602 | 44.06 | 55.12 |
| ALL | 563 | 1,473,167 | 100.00 | 146,100 | 46.46 | 52.37 |

Table 1 also lists the standardised type-token ratio (Std TTR, with a base of 1,000) for each kind of source, from which we may for instance deduce that the Luganda used in the Bible and newspapers (taken from the Internet) is more repetitive than that of the prose found in books (Scans + OCR), especially the self-published ones (e-Transfer).²

In terms of time depth, the corpus covers a period from 1891 to 2011, although, due to the availability of the material, the distribution is not even at all. For the late 1800s (1891) to the early 1900s (1926), we were able to bring together just five files – understandably because the language has a written history dating back to just 120 years only. Fortunately, these files are quite sizeable since they add up to 169,753 words, constituting 11.52% of the entire corpus. We have not yet been successful in obtaining material from the period 1930–1950. In terms of number of files there is a relatively even distribution for the period 1960–1990, for which we gathered 11 files, which add up to 80,672 words and constitute 5.48% of the total corpus. From the year 2000 onwards, we witness a huge increase in the available material, clearly the result of Luganda being used online. It is also

2. In simple terms a Std TTR with a base of 1,000 shows the average number of “new” orthographic words that enter a corpus for every additional 1,000 words added to that corpus. While any base can be chosen (e.g. 500, 1,000, 10,000, etc.), the value itself does imply something about lexical density.

the period when the Luganda Bible of 2003 was uploaded onto the Internet. We gathered about one hundred files for the 2000s, and nearly 450 for just 2010 and 2011. The details are shown in Table 2.

Table 2. Period distribution in the Luganda corpus

| Period | Files | Tokens | % |
|--------|-------|-----------|--------|
| 1890s | 1 | 1,028 | 0.07 |
| 1900s | 2 | 152,258 | 10.34 |
| 1910s | 1 | 9,499 | 0.64 |
| 1920s | 1 | 6,968 | 0.47 |
| 1930s | – | – | – |
| 1940s | – | – | – |
| 1950s | – | – | – |
| 1960s | 2 | 27,434 | 1.86 |
| 1970s | 2 | 577 | 0.04 |
| 1980s | 3 | 50,852 | 3.45 |
| 1990s | 4 | 1,809 | 0.12 |
| 2000s | 101 | 1,063,252 | 72.17 |
| 2010s | 446 | 159,490 | 10.83 |
| ALL | 563 | 1,473,167 | 100.00 |

As far as topics and genres are concerned, the distribution is as shown in Table 3. Despite the fact that there are a massive 400 newspaper articles (out of 563 files, or thus 71.05%), these only constitute 7.05% of all words in the corpus. The largest import comes from religious texts, especially the Bible, which weigh in at 54.88% of the total number of words in the corpus. Cultural and historical texts are also sizeable, at respectively 7.60% and 8.24%.

We queried this Luganda corpus with *WordSmith Tools 6* (Scott 2013), and exported the results of the various searches (i.e. KWIC lines plus source for each line), to a spreadsheet. Although the focus in the present article is on *-sóból-* we also looked at the most important other markers of possibility (i.e. *-yinz-*, *-ándi-*, *-sóbózés-*, *-sóbók-* and *-sóbós-*). For each of the search nodes, we would typically search for the roots of these markers of possibility, surrounded by stars to cater for the inflections. For the very frequent forms, standard sampling techniques were used; for the lesser frequent ones, all instances were considered. Also, frequent deverbative nouns (derived from the markers of possibility) were inserted in the slot of search terms to exclude. After exporting the result of each search to a spreadsheet, the first author used his mother-tongue knowledge to analyse, tag, gloss and annotate all the relevant instances extracted from the corpus. A tag would take the form of Lex for ‘lexical’, P-In DyPo for ‘participant-inherent dynamic possibility’, etc. – to be selected from a drop down attribute list. Glossing

Table 3. Topic/genre distribution in the Luganda corpus

| Topic/Genre | Files | Tokens | % |
|-------------------------|-------|-----------|--------|
| Cultural texts | 6 | 111,910 | 7.60 |
| Environmental documents | 2 | 4,537 | 0.31 |
| Financial texts | 7 | 30,917 | 2.10 |
| Folktales | 1 | 22,614 | 1.54 |
| Health documents | 11 | 17,600 | 1.19 |
| Historical texts | 5 | 121,377 | 8.24 |
| Inspirational materials | 1 | 1,309 | 0.09 |
| Instructional materials | 5 | 62,663 | 4.25 |
| Legal texts | 5 | 70,094 | 4.76 |
| Magazines | 1 | 20,531 | 1.39 |
| Newspaper articles | 400 | 103,891 | 7.05 |
| Novels | 2 | 67,617 | 4.59 |
| Plays | 1 | 5,450 | 0.37 |
| Political documents | 9 | 13,167 | 0.89 |
| Radio broadcasts | 27 | 7,321 | 0.50 |
| Religious texts | 68 | 808,494 | 54.88 |
| Songs | 12 | 3,675 | 0.25 |
| ALL | 563 | 1,473,167 | 100.00 |

would also necessitate the addition of tone marks, given tone is not marked in the standard orthography. Further annotations would typically constitute observations, left in the notes field. Given that each of the 563 source files in the corpus belonged to one particular time period and one particular genre/topic, the source information attached to each KWIC line also allowed us to resort the data along those lines, enabling us to uncover fluctuations and trends through time and to see variations according to text types.

5. On *-sóból-* and the expression of possibility in Luganda

Although there are other markers of possibility in Luganda, we chose to focus our attention on *-sóból-*, being the most prominent of all. Indeed, following a brief sample of the entire corpus for the six most important modal markers of possibility in Luganda, i.e. *-sóból-*, *-yínz-*, *-ándi-*, *-sóbózés-*, *-sóbók-* and *-sóbós-*, *-sóból-* (together with its past tense and perfective aspect form *-sobódde*) turned out to be the most frequent, with a relative frequency of 38% compared to the other five, followed by *-yínz-* with 34%, *-ándi-* with 17% and *-sóbózés-* with 7%. In fifth position is *-sóbók-* with 4%, and lastly *-sóbós-* with 0.16%. See Figure 1 for an overall perspective. The verb forms *-sóbózés-*, *-sóbók-* and *-sóbós-* are derived from the base verb *-sóból-*.

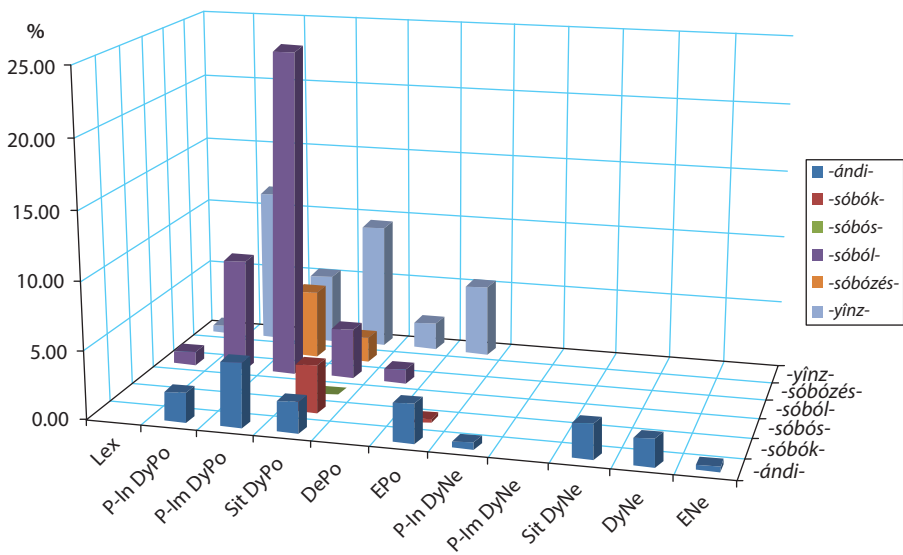


Figure 1. Mapping of the most important modal markers of possibility onto their usage³

Unlike many of the other modal markers, our corpus data also shows that the use of *-sóból-* is restricted to the expression of possibility. It is almost exclusively used as a modal verb in double verb constructions, i.e. followed by the infinitive of the main verb referring to the state of affairs. In only about 3% of the cases it is attested outside double verb constructions.

5.1 *-sóból-* in double verb constructions

In 97% of the corpus attestations, *-sóból-* is followed by an infinitive referring to the main event of the clause. In this capacity, it signals that the state of affairs expressed by the following infinitive is either possible or permitted. It is basically used to express two sub-categories of possibility, i.e. dynamic possibility with its different subdivisions (participant-inherent, participant-imposed, situational) and deontic possibility. In this sub-section, we shall illustrate these uses with examples drawn from different time periods and topics/genres.

3. Abbreviations used: Lex = lexical use; P-In DyPo = participant-inherent dynamic possibility; P-Im DyPo = participant-imposed dynamic possibility; Sit DyPo = situational dynamic possibility; DePo = deontic possibility; EPo = epistemic possibility; P-In DyNe = participant-inherent dynamic necessity; P-Im DyNe = participant-imposed dynamic necessity; Sit DyNe = situational dynamic necessity; DeNe = deontic necessity; ENe = epistemic necessity.

In the Examples (6) to (8) below, *-sóból-* expresses participant-inherent dynamic possibility. In each of the examples, it is immediately followed by the infinitive of a transitive main verb taking an object noun phrase.⁴

- (6) Kye nnee buuzá nti ddala wano mu ggwanga tewali balooya basobola kuwangu ngula musango guno?

kye n-éé-búúz-á nti ddala wa-no mu
 REL₇ SC_{1SG}-REFL-ask-FV that really PP₁₆-DEMA LOC₁₈
li-wánga te-wá-lí ba-lóoya ba-soból-a
 NP₅-country NEG-PP₁₆-be NP₂-lawyer SC₂-POT-FV
ku-wángúl-á mú-sángó gu-no
 NP₁₅-win-FV NP₃-case PP₃-DEMA

“What I am asking myself is that, here in this country, are there no lawyers who can win this case?” (ED110919-Lwaki, Newspapers, 2010s)

- (7) [...] kubanga asobola okuzimba nõkuwanguza KCC [...]

kubánga a-sóból-á ó-kú-zím-b-á né
 because SC₁-POT-FV AUG₁₅-NP₁₅-build-FV and
o-kú-wángúz-á KCC
 AUG₁₅-NP₁₅-win.CAUS-FV KCC

“[...] because he can build and enable KCC to win [...]”⁵

(ED111004-KCC, Newspapers, 2010s)

- (8) Nnina okukkiriza nti tusobola okusobozesa abantu baffe.

n-liná o-kú-kkíríz-a nti tú-soból-á
 SC_{1SG}-have AUG₁₅-NP₁₅-believe-FV that SC_{1PL}-POT-FV
ó-kú-sóbózés-á á-bá-ntu ba-affe
 AUG₁₅-NP₁₅-empower-FV AUG₂-NP₂-person PP₂-POSS_{1PL}

“I believe that we can empower our people.”

(FDC, Political Documents, 2010s)

When marking participant-imposed dynamic possibility, *-sóból-* structurally behaves in the same way as in cases of participant-inherent dynamic possibility. It is also followed by a main verb in the infinitive, which can in turn be followed by an object noun phrase if it is transitive, as in Examples (9) and (10).

4. The interlinear morpheme-by-morpheme glosses are based on the Leipzig Glossing Rules (cf. <http://www.eva.mpg.de/lingua/resources/glossing-rules.php>), with the following category labels: APPL: applicative; AUG_x: augment (aka pre-prefix) of class x; CAUS: causative; CONN_x: connective of class x; DEMa: proximal demonstrative; DEMb: distal demonstrative; FUT: future; FUTf: far future; FUTn: near future; FV: final vowel; LOC_x: locative of class x; NEG: negative; NP_x: noun prefix of class x; OC_x: object concord of class x; PASS: passive; PAST: past; PERF: perfective; PL: plural; POSS_x: possessive of class x; POT: potential; PP_x: pronominal prefix of class x; PRON_x: pronoun of class x; REFL: reflexive; REL_x: relative of class x; SC_x: subject concord of class x; SG: singular; SUBJ: subjunctive.

5. KCC stands for ‘Kampala City Council FC’, a football club.

(9) [...] n'amuutuula ku mubiri asobole okulaba omuzannyo.

ne á-mú-túúl-á ku mu-biri a-sobol-é

and SC₁-OC₁-sit-FV LOC₁₇ NP₃-body SC₁-POT-SUBJ

ó-ku-láb-a o-mu-zannyo

AUG₁₅-NP₁₅-see-FV AUG₃-NP₃-game

“[...] and he sat on top of him to be able to watch the game.”

(Empisa, Cultural Texts, 1900s)

(10) Bobby Williamson yagambye nti wa kuddamu kuzimba bupya omusingi asobole okuwangula Kenya.

bobby williamson a-á-gám-b-ye nti wa kú-ddamu

Bobby Williamson SC₁-PAST-say-PERF that CONN₁ NP₁₅-repeat

ku-zím-b-á bú-pyá ó-mu-síngi a-sobol-é

NP₁₅-build-FV NP₁₄-new AUG₃-NP₃-foundation SC₁-POT-SUBJ

o-kú-wángú-l-á kenya

AUG₁₅-NP₁₅-defeat-FV Kenya

“Bobby Williamson said that he has to rebuild the team so that he can defeat Kenya.”

(ED110905-Tujja, Newspapers, 2010s)

In Example (11), it is followed by a reflexive verb form as the main verb.

(11) [...] basibirira abaagalwa baabwe kondomu basobole okwekuuma.

bá-sib-ír-ír-á á-bá-ágalwá ba-abwe

SC₂-pack-APPL-APPL-FV AUG₂-NP₂-partner PP₂-POSS₂

ø-kóndomu ba-sobol-é o-ku-éé-kuum-a

NP₁₀-condom SC₂-POT-SUBJ AUG₁₅-NP₁₅-REFL-protect-FV

“[...] they pack condoms for their partners so that they can protect themselves.”

(BU110920-Omugaso, Newspapers, 2010s)

In Example (12), *-sóból-* marks situational dynamic possibility. The subject of the clause is inanimate and describes a situation, i.e. the fact of testing for the HIV virus. The modal verb takes a class 7 subject concord here, which is the default impersonal marker in Luganda, as in many other Bantu languages. It could actually also mechanically agree with the grammatical subject, i.e. the infinitive *okukéberwá*, by taking a class 15 subject concord.

(12) Okukeberwa akawuka ka siriimu kisobola okweraliikiriza.

o-ku-kéber-w-á á-ká-wúká ká

AUG₁₅-NP₁₅-test-PASS-FV AUG₁₂-NP₁₂-virus CONN₁₂

síríimu kí-sóból-á ó-kú-éraliikir-iz-a

HIV SC₇-POT-FV AUG₁₅-NP₁₅-stress-CAUS-FV

“Testing for the HIV virus can cause to stress.”

(Kaposi Sarcoma1, Health Documents, 2010s)

In Example (13), finally, *-sóból-* functions as a marker of deontic possibility. The speaker makes a deontic judgement here by referring to a societal norm in Buganda not preventing close contact between a woman and her mother-in-law. From a structural point of view, this example is interesting in that it shows that *-sóból-* and the main verb can be separated by an adverb, in this case *búlúngi nnyo*. The adverb could actually also follow the main verb. None of the orders is preferred.

- (13) [...] omuwala asobola bulungi nnyo okutuula nòmukadde omukazi [...]
o-mu-wála a-sóból-á bú-lúngi nnyo o-ku-túul-a
 AUG₁-NP₁-girl SC₁-POT-FV NP₁₄-good very AUG₁₅-NP₁₅-sit-FV
ne ó-mu-kadde o-mu-kazi
 with AUG₁-NP₁-parent AUG₁-NP₁-woman
 “[...] a girl can comfortably sit with her mother-in-law [...]”
 (ED110905-Empisa, Newspapers, 2010s)

5.2 *-sóból-* outside double verb constructions

In only a tiny minority of corpus attestations (3%), *-sóból-* does not figure in a double verb construction. In such a case, it is mostly found with an anaphoric object concord, as in Example (14), rather than with an object noun phrase as its complement. The only case in our corpus where it takes an object noun phrase is from the 1910s, which we will discuss later. In passing we can also point out that if one uses the full text of the 1009-page modern monolingual Luganda dictionary of Kiingi et al. (2009) as its own corpus, only three instances of *-sóból-* out of a total of 479 attestations figure outside double verb constructions. In all three cases, it does not take an object noun phrase.

The capacity of *-sóból-* to take an object concord immediately preceding the verb root, as in Example (14), indicates that it is basically a transitive verb. As in many other Bantu languages (see for example Riedel 2009), the object concord in Luganda can be considered as an anaphoric pronoun. Unlike the subject concord, which cannot only be in ‘anaphoric agreement’ with an extra-clausal constituent, but also in ‘grammatical agreement’ with a clausal nominal subject (cf. Bresnan & Mchombo (1987) for the distinction between both kinds of agreement), the object concord can never agree with a nominal object in the same clause. Object concords in Luganda are always co-referential with an extra-clausal nominal object, which has usually been mentioned before in the discourse. In Example (14), the class 14 object prefix preceding the root *-sóból-* is co-referential with the pre-posed class 14 object noun phrase *obukúlémbézé* ‘leadership’.

(14) [...] okulaga eggwanga nti obukembeze babusobola.

o-kú-lág-á *é-li-wánga* *nti*
AUG₁₅-NP₁₅-show-FV AUG₅-NP₅-nation that
o-bu-kúlémbézé *bá-bu-sobol-a*
AUG₁₄-NP₁₄-leadership SC₂-OC₁₄-POT-FV

“[...] to show the nation that leadership, they can manage it.”

(BU111019-Okugoba, Newspapers, 2010s)

As transitive verbs are commonly able to do in Luganda, the verb *-sóból-* can also take a reflexive form, at which point it becomes *-éesobol-* “be capable of managing one’s own affairs” (as correctly glossed in Murphy 1972: 504). This reflexive form is mostly used as a subject relative, *eyéésóbóla* (singular) or *abéésóbóla* (plural), which can function as a noun as in Example (15).

(15) [...] okulongoosebwa bisobole okusikiriza abeesobola mu by’ensimbi [...]

o-kú-lóngós-ébw-á *bi-sobol-é* *ó-kú-síkíríz-á*
AUG₁₅-NP₁₅-improve-PASS-FV SC₈-POT-SUBJ AUG₁₅-NP₁₅-attract-FV
á-bá-éé-sóból-a *mu* *byá* *é-n-sím-bi*
AUG₂-PP₂-REFL-POT-FV LOC₁₈ CONN₈ AUG₁₀-NP₁₀-money

“[...] to be improved in order to attract those who are capable of managing their own affairs in terms of money [...]”

(OccupiedBuganda_20070303, Political Documents, 2000s)

The negative counterparts of *eyéésóbóla* and *abéésóbóla*, i.e. *atéésóbóla* and *abátéésóbóla*, behave in the same way, as shown in Example (16), although these forms underwent some degree of lexicalisation to refer specifically to “the handicapped”.

(16) [...] okugolokoka mu kiro ng’olina ateesobola.

o-kú-gólókók-á *mú* *kí-ló* *nga* *o-lí-ngá*
AUG₁₅-NP₁₅-wake.up-FV LOC₁₈ NP₇-night when SC_{2SG}-be-like
a-te-éé-sóból-a
PP₁-NEG-REFL-POT-FV

“[...] to wake up at night feeling like someone who is handicapped.”

(Engero22, Folktales, 1960s)

From what precedes, it has become clear that in recent times, when used outside double verb constructions, the meaning of *-sóból-* is not really distinct from its meaning in double verb constructions. It translates as “to (be able to) manage”, “to be a match for”, “to (be able to) hold one’s own (against)”, “to (be able to) deal with”, as has been done in Luganda dictionaries from the second half of the 20th century, e.g. “be able, succeed, master, manage” (Snnoxall 1967: 290) or “be able to cope with” (Murphy 1972: 504). Hence, also outside double verb constructions, its meaning today can be considered as modal.

However, all the attestations of *-sóból-* outside double verb constructions in the earlier periods of our corpus (i.e. 1890s through 1920s) depict it as having a pre-modal meaning, as in Example (17), where it can be interpreted as meaning “to overcome” or “to defeat”.

- (17) Ntale bwe yalaba nga taasobole Baganda n’adduka n’agenda e Kakoba.
ntále bwe a-a-láb-a nga te-á-sóból-e
 Ntale when SC₁-PAST-see-FV that NEG-SC₁-defeat-FUTn
bagánda ne á-ddúk-á ne á-génd-a e kakóba
 Baganda then SC₁-run-FV and SC₁-go-FV to Kakoba
 “When Ntale saw that he would not defeat the Baganda, he fled and went to
 Kakoba.” (Bassekabaka – 1910s, Historical Texts, 1910s)

This is further evident from several late 19th and early 20th century grammars and dictionaries by French missionaries, where the verb is only used outside double verb constructions with pre-modal meanings. In their oldest Luganda grammar, Livinhac (1885:72) cites *kusobola* and translates it as *dompter, maîtriser*, i.e. “to tame”, “to master”. However, all examples given pertain to its neutro-passive form *-sóbók-*. In a second edition of the same grammar, Livinhac & Denoit (1894: 132) translate *kusobola* as *dompter, maîtriser, en venir à bout, surmonter*, i.e. “to tame”, “to master”, “to overcome”, “to get over”. On the same page, they mention the verb *kusoba* which could be the base verb of *-sóból-*. It is translated as *dépasser le nombre, la mesure, être en plus*, i.e. “exceed in quantity or measure”, “to be additional”, a meaning not totally unconnected to the lexical meaning of *-sóból-*. Murphy (1972:504) defines *-sób-* as “be over and above, lose shape or pattern (e.g. of a mat; be or become pregnant before marriage)”. In the second edition of the Luganda grammar, Livinhac & Denoit (1894: 132) present a few example sentences with *-sóból-*. In Example (18) the neutro-passive derivation *-sóbók-* precedes *-sóból-* taking a class 1 object concord referring to the aforementioned *múntú*.

- (18) Muntu oyo tasoboka, naye nze ndimusobola.
mú-ntú o-yo te-á-sóbók-a nayè nze
 NP₁-person PP₁-DEMB NEG-SC₁-POT-FV but I
n-lí-mu-sobol-a
 SC_{1SG}-FUTf-OC₁-master-FV
Cet homme est intraitable, mais moi je le maîtriserai.
 “That person is inflexible, but I will control him.”
 (Livinhac & Denoit 1894: 132)

In Example (19), the nominal object *mulímó guno* is pre-posed, thus occurring outside the main clause.

(19) Mulimo guno onoogusobola?

mu-límó gu-no o-nóó-gu-sobol-a

NP₃-work PP₃-DEMA SC_{2SG}-FUTn-OC₃-master-FV

Ce travail en viendras-tu à bout?

“This work, will you master it?”

(Livinhac & Denoit 1894: 132)

In Examples (20) and (21), *-sóból-* takes an object concord of class 1 referring to a person previously mentioned in the discourse.

(20) Walumbe, ozze naye, olimusobola?

walúmbe o-zzé naye o-lí-mu-sobol-a

killer SC_{2SG}-come.PERF with.him SC_{2SG}-FUTf-OC₁-master-FV

Mort, tu es venu avec lui, tu le maîtriseras?

“Death, you have come with him, will you tame him?”

(Livinhac & Denoit 1894: 163)

(21) Walumbe bw'oonoomusanga wano, toomusobole; n'ogenda.

walúmbe bwe o-nóó-mu-sáng-á wa-no

killer when SC_{2SG}-FUTn-OC₁-find-FV PP₁₆-DEMA

te-ó-mú-sóból-e ne ó-génd-a

NEG-SC_{2SG}-OC₁-master-FUTn and SC_{2SG}-go-FV

Mort, quand tu le rencontreras ici, tu ne le maîtriseras pas; et tu pars.

“Death, when you will find him here, you will not defeat him; and you go.”

(Livinhac & Denoit 1894: 166)

Looking at one of the oldest dictionaries available (Le Veux 1917:912), *-sóból-* is translated as *pouvoir, réussir, faire aboutir, en venir à bout, maîtriser*, i.e. “to be able, succeed, achieve, overcome, master”, a mixture of modal and pre-modal meanings. Worth noting is the fact that a passive form *sobolebwa* translated as *être maîtrisé*, i.e. “to be mastered”, is provided in this dictionary, which further confirms its transitivity. This form is missing in our entire corpus and is no longer used in current-day Luganda according to the mother-tongue intuition of the first author. Finally, when *-sóból-* is used outside double verb constructions today, in specific contexts involving fighting or war, it could still carry a pre-modal interpretation, thus of defeat or to overcome. However, this usage has become rare.

The pre-modal meaning of *-sóból-* is similar to that of its cognate *-shóbor-* in Kirundi, where it may convey the lexical meanings “to be wealthy” and “to be powerful” outside double verb constructions (Bostoen et al. 2012). To provide additional evidence, we look further at the only noun derived from *-sóból-* in our corpus, i.e. *obúsóbózí*. It can be used to refer to “control, power or authority” as in Example (22), where it refers to God’s control over everything.

(22) Mmanyi nti mazima katonda alina obusobozi ku buli kintu kyonna.

n-manyí nti mazima katonda a-lín-á

SC_{1SG}-know that fact god SC₁-have-FV

ó-bú-sóbózí ku buli ki-ntú ki-onna

AUG₁₄-NP₁₄-control over every NP₇-thing PP₇-all

“I know the fact that God has control over everything.”

(Okugunja, Religious Texts, 2000s)

In Example (23), however, it is used to refer to “wealth”, i.e. that those who have the financial ability can go ahead to hire housemaids.

(23) Abo abalina obusobozi bafune abakozi ab’awaka.

a-bo abá-lín-á o-bu-sobozi

PP₂-DEMB REL₂-have-FV AUG₁₄-NP₁₄-capability

ba-fun-é á-ba-kozi abá á-wá-kâ

SC₂-get-SUBJ AUG₂-NP₂-worker CONN₂ AUG₁₆-NP₁₆-home

“Those who have the capability let them get housemaids.”

(Entanda 7, Magazines, 2000s)

Despite the fact that *obúsóbózí* predominantly appears to have a modal meaning – “ability, capability, competence” (Murphy 1972: 46) or “state of being able to do the required work” (Kiingi 2009: 329) –, the abilities and capabilities are assumed to be inherent in someone with “power” and/or “authority” or who is associated with wealth.

5.3 Is *-sóból-* an auxiliary or not?

Although *-sóból-* mainly occurs in double verb constructions, we have so far avoided calling it an auxiliary. This verb can be seen as an auxiliary if one adopts a rather loose definition as the one proposed by Anderson (2006: 5), i.e. “an element that in combination with a lexical verb forms a mono-clausal verb phrase with some degree of (lexical) semantic bleaching that performs some more or less definable grammatical function”. As shown above, *-sóból-* does most often constitute a mono-clausal verb phrase with a lexical verb and it nowadays only rarely conveys its lexical meaning, not even outside double verb constructions. It has become above all a modal function word whose semantic range is restricted to the domain of possibility. Its grammatical meaning might otherwise – either in Luganda or elsewhere – be expressed by means of an inflectional affix. In a double verb construction, it does not refer to the state of affairs itself, but indicates that the event expressed by the main verb is possible or permitted.

However, like in many other languages of the world, there are few or no formal criteria in Luganda that permit to set *-sóból-* or other auxiliaries apart as a discrete class of verbs. They do not manifest characteristic morphological or syntactic features differentiating them from other types of verbs. Unlike (modal) auxiliaries in, for instance, English or Chinese (Li & Thompson 1989: 172–183, Palmer 1990), *-sóból-* can be inflected and accompanied by different tense and aspect markers; it can occur in non-finite forms such as the infinitive; it can be nominalized; it can take a direct object or an object concord; and it can be modified by intensifiers. Several of the examples that precede and follow illustrate these properties which *-sóból-* shares with common verbs. Nevertheless, there is at least one context in which the morpho-syntactic behaviour of *-sóból-* deviates from that of common verbs, indicating that it is indeed evolving to a true auxiliary with less syntactic autonomy.

In double verb constructions where *-sóból-* expresses dynamic possibility, the main verb can always be omitted and referred to by an anaphoric object marker on the auxiliary. As discussed above, the same is possible when *-sóból-* takes a nominal object, except that the object marker is then co-referential with an object noun phrase. In a double verb construction, the object concord substituting the main verb may belong to class 15, as in Example (24). It then expresses mechanic agreement with the *ku-* noun prefix of the main verb. In many Bantu languages, including Luganda, infinitives take the *ku-* noun prefix of class 15. In Example (24), the infinitive *okúbálá* “to count” is pre-posed.

- (24) [...] kubanga yali amanyi nti okubala akusobola bulungi.
kubánga a-a-lí á-mányi nti
 because SC₁-PAST-be SC₁-know that
o-ku-bal-a a-ku-sóból-á búlúngi
 AUG₁₅-NP₁₅-count-FV SC₁-OC₁₅-POT-FV well
 “[...] because he/she knew that counting, he/she could (do) it well.”
 (Mbayiwa, Novels, 1980s)

The modal verb can also take an object concord of class 7, which is the default impersonal marker in Luganda, to refer to the previously mentioned main verb, as in Example (25) where *-ki-* refers to the verb *okúdduukirira* “to rescue” in the previous sentence. We observed the same pattern in Example (12), where impersonal *-ki-* occurred as a subject concord in association with an infinitive as grammatical subject.

- (25) Omulenzí bw’aba akisobola kirungi nnyo.
o-mú-lénzí bwe a-bá á-ki-sobol-a ki-rúngi nnyó
 AUG₁-NP₁-boy if SC₁-be SC₁-OC₇-POT-FV SC₇-good very
 “If the boy can (do it), then it is very good.”
 (Okwanjula, Cultural Texts, 2000s)

The object concord of class 15 or 7 may be omitted altogether and only remain implied by the context of usage as in Example (26). It is clear that what they could not is expressed by an infinitive form *okúlámúsá* “to greet” in the previous sentence.

- (26) Naye bwe baalaba nga tebaasobole ne bagamba nti [...]
nayé bwe bá-á-lab-á nga te-bá-sóból-e
 but when SC₂-PAST-see-FV that NEG-SC₂-can-FUTn
ne bá-gámb-a nti [...]
 then SC₂-say-FV that [...]
 “But when they saw that they could not, they said that [...]”
 (Empisa, Cultural Texts, 1900s)

The morpho-syntactic behaviour of *-sóból-* as a marker of dynamic modality in a double verb construction is the same as when it takes a nominal object and expresses the same meaning. As discussed above, as other transitive verbs in Luganda, it can only take an object concord when the nominal object with which it is co-referential is extra-clausal. In Example (27), the class 7 object concord *-ki-* is co-referential with the topicalized extra-clausal constituent *ekibínjá kyáyó*. If the latter occurred in its canonical post-verbal object position, the object concord would not be licensed.

- (27) Ekibinja kyayo baali bakifufuggaza lutali luno.
e-ki-bínjá kyá-yó bá-á-lí
 AUG₇-NP₇-group CONN₇-there SC₂-PAST-be
bá-ki-fúfúggáz-á lútálí luno
 SC₂-OC₇-defeat-FV badly
 “The group of that side, they were defeating it badly.”
 (Buwuula, Novels, 2000s)

However, when *-sóból-* conveys deontic possibility, it can never take an object marker substituting the main verb. In the sentence in Example (13), the main verb phrase *okutuúla ne ómukadde omukazi* “to sit with her mother-in-law” could never be represented by an object marker on the auxiliary *-sóból-* expressing deontic possibility. If one does not want to repeat the whole verb phrase, one can only replace it with the verb *okukola* “to do” carrying an impersonal class 7 object marker, i.e. *asóbólá okukikola* “she can do it”. This incompatibility of *-sóból-* with the object concord, when it expresses deontic possibility, could be interpreted in terms of reduced transitivity. When *-sóból-* expresses dynamic modality, the main verb in a double verb construction still behaves as an ordinary nominal object, except that it can be substituted by the impersonal object marker of class 7, while

nominal objects always require an object belonging to the same class. However, when *-sóból-* is a marker of deontic possibility, the main verb cannot be interpreted as a normal nominal object. Just like modal auxiliaries in English, *-sóból-* can no longer take a direct object.

Moreover, the fact that *-sóból-* needs to be followed by a second verb, even if it is the default generic action verb *okukola*, suggests that the syntactic cohesion between *-sóból-* and the main verb is stronger in the case of deontic possibility than in the case of dynamic possibility. It behaves more like a true auxiliary which obligatorily co-occurs with a second verb. Even if an adverb may still intervene between the auxiliary and its auxiliante, as in Example (13), this constraint on object marking suggests that the syntactic status of the auxiliary has become less autonomous. The same tendency was observed in Kirundi (Bostoen et al. 2012). A loss in syntactic independency in the structural domain, i.e. grammaticalisation, goes together with further developed modality in the semantic domain.

What is more, deontic modality is an attitudinal category. Unlike dynamic modality, which only gives information on the state of affairs, deontic modality indicates to what extent someone can commit him/herself to the state of affairs according to certain principles which can be either broad societal norms or rather individual ethical criteria (Nuyts 2006: 5, Nuyts et al. 2010: 17). Although a speaker may describe a deontic assessment held by others, he/she is most typically the one evaluating the state of affairs him/herself. Hence, deontic modality usually has a subjective dimension, or rather intersubjective if the speaker is not relying on a personal opinion, but referring to a general moral principle or the opinion of a wider group of people (Nuyts et al. 2010: 29). The formal process of grammaticalisation from a content word to a grammatical word – an auxiliary in this case – through the loss of syntactic freedom thus also seems to be correlated with ‘subjectification’ as defined by Traugott (1989), i.e. the semantic evolution from ‘objective’ meanings based in the externally described situation, i.e. the outside world, to meanings which “tend to become based in the speaker’s subjective belief state/attitude toward the proposition” (Traugott 1989: 35). Luganda is not unique in this regard. Modal verbs have been treated as prominent instances of the interaction between grammaticalisation and subjectification in the world’s languages (Narrog 2010: 387–392).

5.4 Discussion and comparison with Kirundi

The examples above illustrate the semantic range of *-sóból-* as attested in our corpus. In 1.5 million words it occurs 1,647 times, or thus on average 11 times for every 10,000 words. The frequency distribution of the different types of meaning it covers is represented in Figure 2. The verb is mostly used in double verb constructions and most prominently in the expression of dynamic possibility. We counted 362 cases where it is used to express participant-inherent dynamic possibility (22%) and 1,045 cases where it is used to express participant-imposed dynamic possibility (63%). There were 159 cases (10%) in which it is used to express situational dynamic possibility. In 95% of the cases (1,566 out of 1,647), *-sóból-* thus conveys dynamic possibility. Lastly, in only 40 cases (2%) was the verb found to express deontic possibility. This is even less than the marginal occurrence of *-sóból-* outside double verb constructions, of which we found only 41 attestations (3%) in our corpus (cf. Lex in Figure 2).

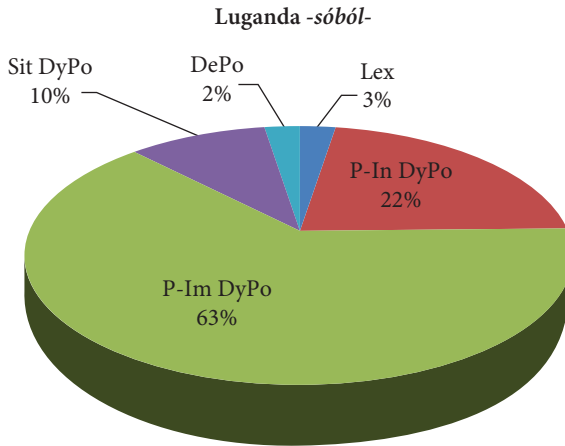


Figure 2. Semantic range of Luganda *-sóból-*

From a comparative point of view, it is interesting to note that the modal uses of *-sóból-* in Luganda are less diversified compared to those of its cognate *-shóbor-* in Kirundi. Whereas the Kirundi *-shóbor-* “covers all potential sub-categories of the semantic domain of modality” (Bostoen et al. 2012: 13), including epistemic possibility, as seen in Figure 3, *-sóból-* mainly covers dynamic possibility and, to a much lesser extent, deontic possibility. It is not involved in the expression of epistemic possibility at all. Although the pre-modal meaning of both cognate verbs is similar and has become relic in both languages, their semantic evolution

within the modal domain of possibility clearly did not follow the same pace. In the Kirundi corpus, which contains mostly recent material, *-shóbor-* is also more frequent than *-sóból-* is in the Luganda corpus: it occurs on average 21 times for every 10,000 words.

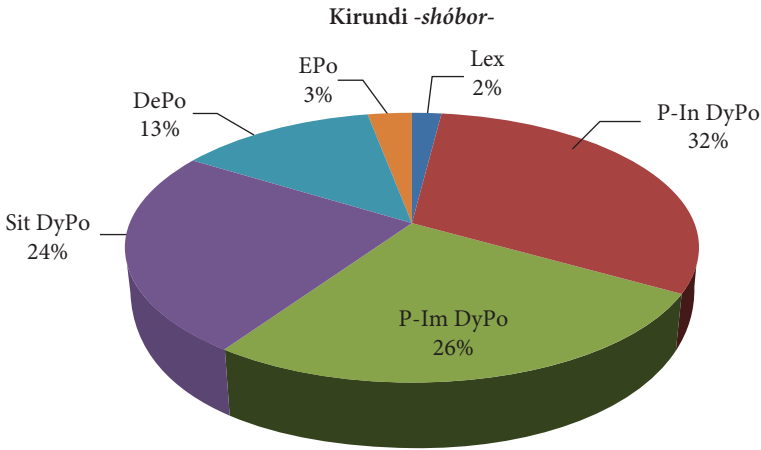


Figure 3. Semantic range of Kirundi *-shóbor-* (adapted from Bostoen et al. 2012)

Given that the Luganda modal marker *-sóból-* evolved differently from its Kirundi equivalent *-shóbór-*, both in the semantic and structural domain, we will now examine more closely how *-sóból-* evolved over the last century within Luganda. To do this, we submit *-sóból-* to a diachronic corpus analysis.

6. Diachronic corpus analysis of the use of *-sóból-*

We selected three time periods from our corpus: “the 1900s” (i.e. 1890–1920), “the 1960s” (i.e. 1960–1969) and “the 2010s” (i.e. 2010 and 2011). The sizes of these three sub-corpora are 169,753 words, 27,434 words and 159,490 words respectively. Unfortunately, the size of the middle section is different from the two extremes, but at least the latter two are comparable in size.

6.1 Distribution

One firstly notices that the use of *-sóból-* grows exponentially across these three time periods, as illustrated in Figure 4.

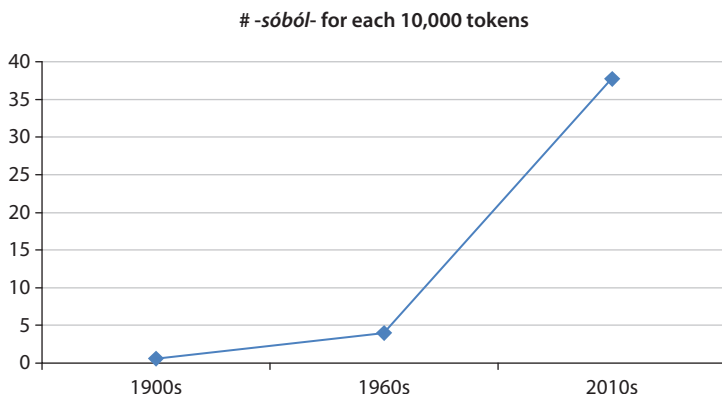


Figure 4. Overall use of *-sóból-* across time

In the 1900s *-sóból-* is rare, with only 0.59 instances for every 10,000 words. In absolute terms this corresponds to just 10 attestations in the 1900s sub-corpus (of 169,753 words): five times as a full verb, two times as participant-inherent dynamic possibility, and three times as participant-imposed dynamic possibility. The verb was not found to express any other type of possibility in the 1900s.

In the 1960s *-sóból-* is more frequent, with 4.01 instances for every 10,000 words. In absolute terms this corresponds to 11 attestations in the 1960s sub-corpus (of 27,434 words): once only as a full verb, and four times each as participant-inherent and participant-imposed dynamic possibility. Compared to the 1900s, a new type also appears, in the form of two attestations as situational dynamic possibility.

In the 2010s the use of *-sóból-* jumps to 37.81 instances for every 10,000 words. In absolute terms this corresponds to a massive 603 attestations in the 2010s sub-corpus (of 159,490 words): 13 times as an autonomous verb, 121 times as participant-inherent dynamic possibility, 285 times as participant-imposed dynamic possibility, and 171 times as situational dynamic possibility. Compared to the 1960s, yet another type also appears, in the form of 13 attestations as deontic possibility.

With these details, Figure 4 can be redrawn as Figure 5, now taking the various types of possibility into account.

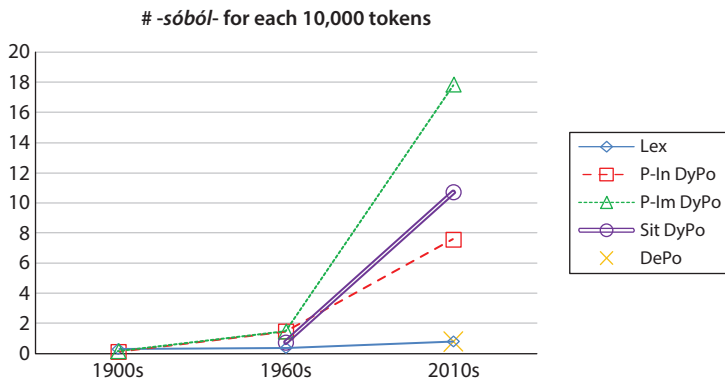


Figure 5. Meanings of *-sóból-* across time

6.2 Discussion

From Figures 4 and 5 we see that the current-day most prominent modal marker of possibility in Luganda, *-sóból-*, used to be much less prominent in double verb constructions in earlier times. If one nonetheless looks at the percentage distribution of the various modal uses across the three sampled periods, Figure 6 is obtained.

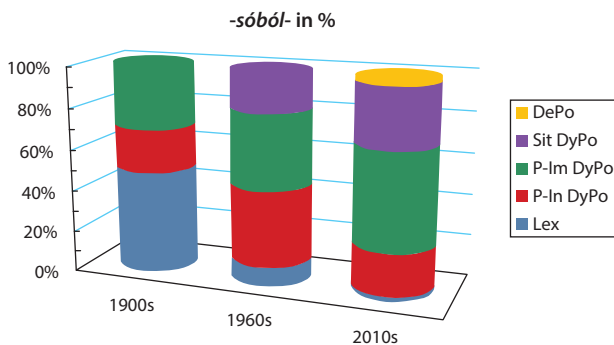


Figure 6. Diachronic semantic transition of *-sóból-*

From Figure 6 one could be tempted to conclude that while the use of *-sóból-* as a full verb steadily decreases compared to the other uses, the use as participant-inherent dynamic possibility remains stable, and the use as participant-imposed dynamic possibility concurrently increases. Also, halfway situational dynamic possibility seems to be added as a modal use of *-sóból-*, a use which increases

further with time, and most recently deontic possibility also seems to be introduced. These would be nice “trends” indeed, but unfortunately a two-tailed Fisher’s exact test reveals that all comparisons involving the middle column (“the 1960s”) are not statistically relevant. Comparing the 2010s with the 1900s, however, does indicate that the decrease in the autonomous use of *-sóból-* is statistically relevant, as is the appearance of situational dynamic possibility.

6.3 Countering corpus imbalance

Although the trends seen in Figure 6 appear to be plausible, the discussion in Section 6.2 invites us to look closer at the composition of the 1900s sub-corpus, which contains just five texts (cf. Table 2). All five texts in that corpus belong to two genres only, historical and cultural, and were all written by the same author, Apollo Kagga (who was also one of the first writing original, i.e. non-translated, texts in Luganda). Two extra texts are thus in order, widening both the authorship and the text genres.⁶ New material from the 1900s could as yet not be pinpointed, but quite early material, from the 1930s and 1940s, could.

In other words, in order to avoid author and genre bias, there was a need to collect extra evidence to support our findings. To this end, two sets of texts were identified. First, we compared two editions of the Luganda Bible (none of which was used when focusing on the diachronic use of *-sóból-* as described above). An edition from 1937 (itself based on an edition from 1896, cf. Mojola 2000) was compared with the latest edition, from 2003. A total of 18 instances were found of the use of *-sóból-* in the earlier (1937) version, 17 times as an autonomous verb and only once in a double verb construction; all of which also appear, in the same way, in the recent (2003) version. However, the recent version also contains two more uses of *-sóból-*, namely in John 9:30, as shown in Example (28), and in Ephesians 6:13, as shown in Example (29). The form *obútámányá* in (28a) is a negative infinitive, which takes a class 14 noun prefix instead of the class 15 noun prefix used in the affirmative, at which point the negative marker is *-ta-* rather than *-te-*.

- (28) a. Kino kitalo mmwe obutamanya gy’ava omuntu eyasobola okunzibula
amaaso!
ki-no ki-táló mmwe o-bú-tá-mány-á gye
PP₇-DEMA NP₇-pity PRON_{2PL} AUG₁₄-NP₁₄-NEG-know-FV where

6. Trying to circumvent the (potential) problem of “a few texts in a small corpus” is not unique to Bantu language corpora, of course, as is for instance shown by the recent research of Ros-souw & Van Rooy (2012: 22, note 9), on diachronic changes in modality in White South African English.

a-v-á *o-mú-ntú* *á-á-á-sóbol-a*
 SC₁-come.from-FV AUG₁-NP₁-person REL₁-SC₁-PAST-able-FV
o-kú-n-zibúl-á *á-má-aso*
 AUG₁₅-NP₁₅-OC_{1SG}-open-FV AUG₆-NP₆-eye

“It is unfortunate if you don’t know where he comes from, the person who was able to open my eyes!” [-*sóbol-* expresses P-In DyPo]

(Bible in Luganda 2003)

- b. Kino kitalo oba mmwe temumanyi gy’ava era yanzibula amaaso.

ki-no *ki-táló* *obá* *mmwe* *te-mú-mányí* *gye*
 PP₇-DEMA NP₇-pity if PRON_{2PL} NEG-SC_{2PL}-know where
a-v-á *erá* *a-á-n-zibúl-a* *a-má-aso*
 SC₁-come.from-FV and SC₁-PAST-OC_{1SG}-open-FV AUG₆-NP₆-eye

“It is unfortunate if you don’t know where he comes from, and he opened my eyes.” [-*sóbol-* is absent]

(Bible in Luganda 1937)

- (29) a. [...] mulyoke muyinzenga okuguma ku lunaku olubi era bwe mulimala okukola byonna musobole okuyimirira.

mu-lyok-é *mú-yínz-é-nga* *o-kú-gúm-á*
 SC_{2PL}-AUX-SUBJ SC_{2PL}-may-SUBJ-always AUG₁₅-NP₁₅-be.firm-FV
kú *lú-náku* *o-lu-bí* *erá* *bwe* *mú-li-mál-á*
 LOC₁₇ NP₁₁-day AUG₁₁-NP₁₁-bad and when SC_{2PL}-FUTf-finish-FV
ó-kú-kól-a *bi-onná* *mu-sobol-é*
 AUG₁₅-NP₁₅-do-FV PP₈-all SC_{2PL}-be.able-SUBJ
o-ku-yimirir-a

AUG₁₅-NP₁₅-stand-FV

“[...] so that⁷ you may take a stand on an evil day, and after you have done everything, you will be able to stand.” [-*sóbol-* expresses P-Im DyPo]

(Bible in Luganda 2003)

- b. (Identical as (29a), except for the omission of *mu-sobol-é*.)

“[...] so that you may take a stand on an evil day, and after you have done everything, to stand.” [-*sóbol-* is absent]

(Bible in Luganda 1937)

Even though the Luganda Bible is a translated text, the recent versions seen in (28a) and (29a) cannot be said to be “better” translations than the earlier ones seen in (28b) and (29b). Today’s translators simply must have felt that more natural, modern Luganda was rendered through the addition of a modal use of *-sóbol-* in these contexts. If anything, then, the use of *-sóbol-* is indeed on the rise, with the new uses tending towards modal uses. These counts are too low to make them statistically relevant, however.

7. *-lyok-* is an auxiliary verb which may introduce, among other things, a subordinate clause, and in such cases it corresponds to the English conjunction “so that” (cf. Murphy 1972: 314).

In addition to the electronic Luganda corpus already built, we also have in our possession a small private collection of handwritten letters from the early and middle 20th century. These letters, a genre not yet covered in our corpus, await digitization, but reading through them we were struck by what we could call their “archaic language use”. This both in terms of the words used (lexicon) and the meanings carried by those words (semantics). One type is relevant here, namely the (over) use of *-yinz-* where today we would use *-soból-*. Here, in Example (30), an instance taken from a letter written in 1944:

- (30) Saayinza kukuddamu mangu.
sí-á-yinz-a *ku-kú-ddamu* *mangu*
 NEG_{1SG}-PAST-POT-FV NP₁₅-OC_{2SG}-reply quickly
 “I was not able to reply you quickly.”

Today the usage in Example (31) would apply:

- (31) Saasobola kukuddamu mangu.
sí-á-sobol-a *ku-kú-ddamu* *mangu*
 NEG_{1SG}-PAST-POT-FV NP₁₅-OC_{2SG}-reply quickly
 “I was not able to reply you quickly.”

Uses of *-yinz-* such as the one in Example (30) suggest that more diachronic onomasiological research is needed on other modal verbs in Luganda in order to put the diachronic semasiological evolution of *-soból-* in perspective.

7. Conclusions and further work

Although the field of Bantu corpus linguistics is still young, we have seen that it now fully moved from studies that are corpus-based (since 1999) to those that are corpus-driven (since 2010). Methodologically, the current research proved the feasibility of another first for Bantu corpus linguistics, namely the application of a corpus-driven approach on a diachronic corpus (here spanning twelve decades). Going forward, more corpus-driven diachronic Bantu linguistic studies may be envisaged, but the bottleneck will in each case be the sparseness of early data. The time depth for the Bantu languages will typically not be much deeper than a good century, given extant written and oral material is only this old.⁸ The present research has shown, however, that a century may indeed be enough to uncover dia-

8. The only two obvious exceptions are Kikongo and Swahili, for which (translated) texts exist which date back to the early and mid-seventeenth century.

chronic trends in Bantu, and thus to derive hypotheses directly from the written evidence. We also saw that such studies necessitate (near-)native speaker knowledge, in order to make sense of the fine-grained nuances found in the corpus.

From our diachronic corpus analysis, together with the analysis of additional evidence (i.e. the 1937 edition of the Luganda Bible compared to a recent one from 2003, and a letter written in 1944), we have come to note that the use of *-sóból-* was infrequent during the early 20th century. During that period the verb was equally used outside and inside double verb constructions. Its modal uses were limited to the expression of participant-inherent dynamic possibility and participant-imposed dynamic possibility. In the 1960s, it became more prominent and its uses in double verb constructions also became more diversified, with situational dynamic possibility entering its semantic range. Its use as a marker of participant-inherent dynamic possibility in double verb constructions increased from about 20% to 40%. However, in the 2010s, the verb is increasingly involved in modal double verb constructions, with now also deontic possibility hesitatingly entering its semantic range. Its use to mark participant-inherent dynamic possibility reduces to about 10%, while participant-imposed dynamic possibility plus situational dynamic possibility and deontic possibility – which together constitute participant-external possibility in Van der Auwera & Plungian (1998) – rise to about 90%. The independent use of *-sóból-* is disappearing, but still existent in the language. All this leads us to state that *-sóból-* is on the path of grammaticalisation from a full verb to an auxiliary and this grammaticalisation also involves semantic generalisation, in that the more its modal uses in the semantic domain of possibility have become diversified, the more its lexical use has become insignificant. Nevertheless, unlike its cognate *-shóbor-* in Kirundi, *-sóból-* is not (yet) associated with epistemic possibility in Luganda.

Moreover, the recent development of the more attitudinal or subjective modal meaning of deontic possibility involves a stronger syntactic cohesion between *-sóból-* and the main verb within the mono-clausal double verb construction. This is not the case when *-sóból-* conveys dynamic possibility, an objective meaning pertaining to the outside world. Just like in Kirundi, the semantic shift from the more objective grammatical meaning of dynamic possibility to the more subjective grammatical meaning of deontic possibility goes hand in hand with a reduced syntactic independency of the auxiliary. As is the case in many other languages of the world, subjectification in the semantic domain of modality happens to be correlated in Luganda with grammaticalisation in the structural domain.

The data on lexical and auxiliary uses of *-sóból-* fit in the semantic map of modality proposed by Van der Auwera & Plungian (1998), especially in respect to ‘diachrony’, which is a central aspect of our analysis. For their semantic map, they

state that “most if not all of the modal notions are directly accessible from pre-modal notions to participant-internal modality, participant-external modality [...] and epistemic modality” (Van der Auwera & Plungian 1998: 111). Numerous examples of the premodal meaning of *-sóból-* were not directly retrievable from our original diachronic text corpus. However, the late 19th century Luganda descriptions by French missionaries do provide extra evidence for the verb’s pre-modal meaning. Like its cognate *-shóbor-* in Kirundi, the lexical meaning of *-sóból-* is to be situated in the realms of power, i.e. “to dominate, master, control, overcome”. Although the independent use of *-sóból-* persists throughout the different time periods of our corpus, it becomes increasingly less prominent.

Needless to say, these conclusions could not have been arrived at without the use of a corpus. A 1.5-million-word Luganda corpus is a good start, but with the encompassing study of the full range of modality in Luganda in mind, this study of a first aspect also indicated that we will need a larger corpus, with more data for especially the earlier periods, in order to draw stronger conclusions.

References

- Anderson, G. D. S. 2006. *Auxiliary Verb Constructions*. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199280315.001.0001
- Ashton, E. O., Mulira, E. M. K., Ndawula, E. G. M. & Tucker, A. N. 1954. *A Luganda Grammar*. London: Longmans, Green & Co.
- Bentley, W. H. 1887. *Dictionary and Grammar of the Kongo Language, as Spoken at San Salvador, the Ancient Capital of the Old Kongo Empire, West Africa*. London: Baptist Missionary Society and Trübner & Co.
- Bible in Luganda. 1937. *Ekitabo Ekitukuvu eyka Katonda kye Bayita Baibuli ye Ndagaano Enkadde Nempya awamu Nebijulira*. London: British and Foreign Bible Society.
- Bible in Luganda. 2003. Online. *Ekitabo Ekitukuvu: Baibuli y’Oluganda*. Available at: <http://www.lugandaministry.org/> (accessed October 2013).
- Bostoen, K., Mberamihigo, F. & de Schryver, G.-M. 2012. “Grammaticalization and subjectification in the semantic domain of possibility in Kirundi (Bantu, JD62)”. *Africana Linguistica*, 18, 5–40.
- Bresnan, J. & Mchombo, S. 1987. “Topic, pronoun and agreement in Chichewa”. *Language*, 63 (4), 741–782. DOI: 10.2307/415717
- De Pauw, G., de Schryver, G.-M., Pretorius, L. & Levin, L. 2011. “Introduction to the special issue on African Language Technology”. *Language Resources and Evaluation*, 45 (3), 263–269. DOI: 10.1007/s10579-011-9157-9
- De Pauw, G., de Schryver, G.-M. & van de Loo, L. 2012. “Resource-light Bantu part-of-speech tagging”. In G. De Pauw, G.-M. de Schryver, M. L. Forcada, K. Sarasola, F. M. Tyers & P. W. Wagacha (Eds.), *Proceedings of the Workshop on Language Technology for Normalisation of Less-Resourced Languages (SALTMIL 8 – AFLAT 2012)*. Istanbul: European Language Resources Association, 85–92.

- De Schryver, G.-M. 1999. *Cilubà Phonetics, Proposals for a 'Corpus-based Phonetics from Below'-Approach*. 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. 2010. "Revolutionizing Bantu lexicography: A Zulu case study". *Lexikos*, 20, 161–201. DOI: 10.4314/lex.v20i1.62689
- De Schryver, G.-M. & Gauton, R. 2002. "The Zulu locative prefix *ku-* revisited: A corpus-based approach". *Southern African Linguistics and Applied Language Studies*, 20 (4), 201–220. DOI: 10.2989/16073610209486311
- De Schryver, G.-M. & Nabirye, M. 2010. "A quantitative analysis of the morphology, morphophonology and semantic import of the Lusoga noun". *Africana Linguistica*, 16, 97–153.
- De Schryver, G.-M. & Prinsloo, D. J. 2000. "The compilation of electronic corpora, with special reference to the African languages". *Southern African Linguistics and Applied Language Studies*, 18 (1–4), 89–106.
- De Schryver, G.-M. & Taljard, E. 2006. "Locative trigrams in Northern Sotho, preceded by analyses of formative bigrams". *Linguistics*, 44 (1), 135–193.
- Devos, M. 2008. "The expression of modality in Shangaci". *Africana Linguistica*, 14, 3–35.
- Fourie, D. J. 1989. "Modality in IsiZulu". *Logos*, 9, 45–56.
- Fourie, D. J. 1991. "On the modal system of Ndonga". *South African Journal of African Languages*, 11 (3), 88–91.
- Gauton, R., de Schryver, G.-M. & Mohlala, L. 2004. "A corpus-based investigation of the Zulu nominal suffix *-kazi*: A preliminary study". In A. Akinlabi & O. Adesola (Eds.), *Proceedings of the 4th World Congress of African Linguistics, New Brunswick 2003*. Cologne: Rüdiger Köppe, 373–380.
- Hanks, P. 2013. *Lexical Analysis: Norms and Exploitations*. Cambridge, MA: The MIT Press. DOI: 10.7551/mitpress/9780262018579.001.0001
- Himmelman, N. P. 1998. "Documentary and descriptive linguistics". *Linguistics*, 36 (1), 161–195. DOI: 10.1515/ling.1998.36.1.161
- Himmelman, N. P. 2006. "Language documentation: What is it and what is it good for?" In J. Gippert, N. P. Himmelman & U. Mosel (Eds.), *Essentials of Language Documentation*. Berlin: Mouton de Gruyter, 1–30. DOI: 10.1515/9783110197730
- Kiingi, K. B., Kawalya, D., Nkonge, D. K., Kasiriivuvu, D., Ssentongo I. & Kiriggwajjo, A. 2009. *Enkuluze ya Oluganda eya e Makerere* (2nd ed.). Kampala: Fountain Publishers.
- Le Veux, H. 1917. *Premier Essai de Vocabulaire Luganda – Français d'après l'Ordre Étymologique*. Maison-Carrée (Algiers): Imprimerie des Missionnaires d'Afrique (Pères Blancs).
- Lewis, M. P., Simons, G. F. & Fennig, C. D. (Eds.) 2013. Online. *Ethnologue: Languages of the World, Seventeenth edition*. Available at: <http://www.ethnologue.com/> (accessed October 2013).
- Li, C. N. & Thompson, S. A. 1989. *Mandarin Chinese: A Functional Reference Grammar*. Berkeley, CA: University of California Press.
- Livinhac, L. 1885. *Essai de Grammaire Ruganda*. Paris: Imprimerie F. Levé.
- Livinhac, L. & Denoit, C. 1894. *Manuel de Langue Luganda: Comprenant la Grammaire et un Recueil de Contes et de Légendes* (2nd ed.). Einsiedeln: Benziger & Co.
- Louwrens, L. J. 1990. "Mood and modality in Northern Sotho". *South African Journal of African Languages*, 10 (1), 10–17.

- Mojola, A.O. 2000. "100 years of the Luganda Bible (1896–1996): A general survey". In G. O. West & M. W. Dube (Eds.), *The Bible in Africa: Transactions, Trajectories, and Trends*. Brill: Leiden, 524–537.
- Murphy, J.D. 1972. *Luganda – English Dictionary*. Washington, DC: The Catholic University of America Press.
- Narrog, H. 2010. "(Inter)subjectification in the domain of modality and mood: Concepts and cross-linguistic realities". In K. Davidse, L. Vandelanotte & H. Cuyckens (Eds.), *Subjectification, Intersubjectification and Grammaticalization*. Berlin: Mouton de Gruyter, 385–430. DOI: 10.1515/9783110226102.4.385
- Nash, J.A. 1992. "Aspects of Ruwund Grammar". Unpublished PhD dissertation, University of Illinois at Urbana-Champaign, Urbana-Champaign, IL.
- Ngcobo, M.N. 2010. "Zulu noun classes revisited: A spoken corpus-based approach". *South African Journal of African Languages*, 30 (1), 11–21.
- Nuyts, J. 2005. "The modal confusion: On terminology and concepts behind it." In A. Klinge & H. H. Müller (Eds.), *Modality: Studies in Form and Function*. London: Equinox, 5–38.
- Nuyts, J. 2006. "Modality: Overview and linguistic issues". In W. Frawley, E. Eschenroeder, S. Mills & T. Nguyen (Eds.), *The Expression of Modality*. Berlin: Mouton de Gruyter, 1–26.
- Nuyts, J., Byloo, P. & Diepeveen, J. 2010. "On deontic modality, directivity, and mood: The case of Dutch *mogen* and *moeten*". *Journal of Pragmatics*, 42 (1), 16–34. DOI: 10.1016/j.pragma.2009.05.012
- Palmer, F.R. 1990. *Modality and the English Modals*. London: Longman.
- Palmer, F.R. 2001. *Mood and Modality* (2nd ed.). Cambridge: Cambridge University Press. DOI: 10.1017/CBO9781139167178
- Prinsloo, D.J. & de Schryver, G.-M. 2001. "Corpus applications for the African languages, with special reference to research, teaching, learning and software". *Southern African Linguistics and Applied Language Studies*, 19 (1–2): 111–131. DOI: 10.2989/16073610109486279
- Riedel, K. 2009. *The Syntax of Object Marking in Sambia. A comparative Bantu Perspective*. Utrecht: Landelijke Onderzoeksschool Taalwetenschap.
- Rossouw, R. & van Rooy, B. 2012. "Diachronic changes in modality in South African English". *English World-Wide*, 33 (1), 1–26. DOI: 10.1075/eww.33.1.01ros
- Scannell, K.P. 2003–2013. Online. *Corpus Building for Minority Languages*. Available at: <http://borel.slu.edu/crubadan/index.html> (accessed October 2013).
- Schoenbrun, D.L. 1994. "Great Lakes Bantu: Classification and settlement chronology". *Sprache und Geschichte in Afrika*, 15, 91–152.
- Scott, M. 1996–2013. Online. *WordSmith Tools. Version 6*. Available at: <http://www.lexically.net/wordsmith/> (accessed October 2013).
- Snoxall, R.A. 1967. *Luganda – English Dictionary*. London: Oxford University Press.
- Taljar, 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.
- Taljar, E. 2012. "Corpus-based language teaching: An African language perspective". *Southern African Linguistics and Applied Language Studies*, 30 (3), 377–393. DOI: 10.2989/16073614.2012.739318
- Tognini-Bonelli, E. 2001. *Corpus Linguistics at Work*. Amsterdam: John Benjamins. DOI: 10.1075/scl.6
- Toscano, M. & Sewangi, S. 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.

- Traugott, E. C. 1989. "On the rise of epistemic meanings in English: An example of subjectification in semantic change". *Language*, 65, 31–55. DOI: 10.2307/414841
- Van der Auwera, J. & Plungian, V.A. 1998. "Modality's semantic map". *Linguistic Typology*, 2, 79–124. DOI: 10.1515/lity.1998.2.1.79
- Zawada, B. & Ngcobo, M.N. 2008. "A cognitive and corpus-linguistic re-analysis of the acquisition of the Zulu noun class system". *Language Matters*, 39 (2), 316–331. DOI: 10.1080/10228190802579718

Authors' addresses

Deo Kawalya
Department of Languages and Cultures
Ghent University
Rozier 44, 9000 Ghent
Belgium
deo.kawalya@UGent.be

Koen Bostoën
Department of Languages and Cultures
Ghent University
Rozier 44, 9000 Ghent
Belgium
koen.bostoën@UGent.be

Gilles-Maurice de Schryver
Department of Languages and Cultures
Ghent University
Rozier 44, 9000 Ghent
Belgium
gillesmaurice.deschryver@UGent.be