A diachronic corpus-driven study of the expression of possibility in Luganda (Bantu, JE15)

Deo Kawalya[i], Koen Bostoen[ii], and Gilles-Maurice de Schryver[ii,iii]
[i] Makerere University | [ii] Ghent University | [iii] University of Pretoria

This article employs a 4-million-word diachronic corpus to examine how the expression of possibility has evolved in Luganda since the 1890s to the present, by focusing on the language's three main potential markers *-yînz-*, *-sóból-* and *-andi-*, and their historical interaction. It is shown that while the auxiliary *-yînz-* originally covered the whole modal subdomain of possibility, the auxiliary *-sóból-* has steadily taken over the more objective categories of dynamic possibility. Currently, *-yînz-* first and foremost conveys deontic and epistemic possibility. It still prevails in these more subjective modal categories even though the prefix *-andi-*, a conditional marker in origin, has started to express epistemic possibility since the 1940s, and *-sóból-* deontic possibility since the 1970s. More generally, this article demonstrates the potential of corpus linguistics for the study of diachronic semantics beyond language comparison. This is an important achievement in Bantu linguistics, where written language data tend to be young.

Keywords: possibility, diachronic semantics, subjectification, Luganda

1. Introduction

In descriptions of African languages, the study of meaning and semantic change is still very much in its infancy, especially compared to more formal aspects of language. Within Bantu linguistics, the development of diachronic semantics is hampered by the lack of ancient language records. In Bantu-speaking Africa, not to say most of Sub-Saharan Africa, studying language change is thus hardly ever possible on the basis of historical evidence. To reconstruct the evolution of linguistic form, linguists therefore commonly rely on the Comparative Method, which has a long-standing tradition within Bantu historical linguistics. However, the recon-

struction of semantic change through language comparison is much less advanced (Bastin, 1985; Schadeberg, 2002; Bostoen & Bastin, 2016). As Fleisch (2008: 67) pointed out, "there is hardly any agreement among Bantu comparativists on the important question of how to deal with diachronic aspects of word meaning".

Nevertheless, over the past few decades, a new pathway for the historical study of meaning has emerged beyond the limits of Bantu comparative linguistics, by means of the use of text corpora to examine semantic change within individual languages. Even if most Bantu language corpora have a limited time-depth, typically from the late 19th century onwards at the earliest, and tend to be rudimentary in terms of mark-up or tagging, they have proven to be instrumental for the study of meaning and semantic change (e.g. de Schryver & Nabirye, 2010; Tramutoli, 2015). With a 4-million-word diachronic corpus now available for Luganda, it is therefore our aim, in this article, to examine the evolution of the expression of possibility in Luganda since the 1890s up to the present by focusing on the language's three main potential markers *-yînz-*, *-sóból-* and *-andi-* and their historical interaction. In the process, an overall description and distribution of these markers is also provided.

2. Modality in Bantu languages

Corpus studies of modality are gaining traction in the field of Bantu studies. Especially the modal systems of two Great Lakes Bantu languages, Luganda (JE15) and Kirundi (JD62), have been the subject of detailed corpus studies (see Bostoen et al., 2012; Kawalya et al., 2014; Mberamihigo, 2014; Mberamihigo et al., 2016; Kawalya et al., 2018a). This exploratory research has demonstrated that the corpus methodology is as insightful for the study of modality in Bantu, where it is still nascent (Devos, 2008: 4; Bostoen et al., 2012: 6; Kawalya et al., 2014: 63), as it is in other languages of the world (see van der Auwera & Diewald, 2012).

Modality has received many different definitions and categorisations in the literature. In this paper, we adopt the one offered by Jan Nuyts (Nuyts, 2006).[1] It is also the one which was relied on for the earlier studies of modal markers in Kirundi (Bostoen et al., 2012; Mberamihigo, 2014; Mberamihigo et al., 2016) and in Luganda (Kawalya et al., 2014; Kawalya et al., 2018a, b, 2019). In this approach, modality is a semantic subdomain within the wider tense-aspect-modality domain, involving the concepts of possibility and necessity. Nuyts distinguishes between three categories of modality: dynamic, deontic and epistemic modality.

1. For a brief overview of definitions and typologies for modality as found in the literature, please see the online supplementary material (Appendix B).

'Dynamic modality' refers to capacities/abilities or necessities ascribed to the first-argument participant in the state of affairs. This is further subdivided into (i) 'participant-inherent dynamic modality', to refer to abilities or needs that are fully inherent to the first-argument participant; (ii) 'participant-imposed dynamic modality', which covers abilities or needs which are determined by external factors; and (iii) 'situational dynamic modality', covering cases which characterise a potential or a necessity inherent in the situation described in the clause as a whole (Nuyts, 2006: 3–4). Traditionally defined in terms of permission and obligation, 'deontic modality' is treated in more general terms by Nuyts (2006: 4) as "an indication of the degree of moral desirability of the state of affairs expressed in the utterance." The third type, 'epistemic modality', involves an estimation of the chances or likelihood that the state of affairs expressed in the clause applies in the world (Nuyts, 2006: 6).

As for Luganda, the diachronic study of modality has been initiated in Kawalya et al. (2014) and Kawalya et al. (2018a) through corpus-driven analyses of the auxiliary *-sóból-* and the verbal prefix *-andi-* respectively, both involved in the expression of possibility. The two studies succeeded in collecting unique empirical evidence for change in meaning "at work" since the 1890s. At the same time, they led to the conclusion that a detailed corpus-based study of the modal auxiliary *-yînz-* is also necessary, in order to fully grasp the diachronic evolution of the semantic domain of possibility in Luganda. Kawalya et al. (2014) have already shown this verb to be one of the most frequent markers of possibility along with *-sóból-*. That is why we primarily focus here on the corpus-driven analysis of *-yînz-* in order to subsequently provide a global account of semantic change within the modal subdomain of possibility over the past 13 decades.

3. The Luganda corpus and analytical methods

A large and representative corpus is important in any study if statements derived from its analysis are to be applicable to the language as a whole (Tognini-Bonelli, 2001: 57). In this respect, efforts were made to augment the earlier general 1.5-million-word Luganda corpus (Kawalya et al., 2014), especially focusing on earlier periods that were underrepresented or not represented at all. The corpus used for this study consists of both written and (transcribed) oral material in Luganda and contains 4,053,739 running words or tokens and 337,965 distinct

words or types. It is organised into 13 decades (1890s to 2010s) and 18 topics/genres, as shown in Tables 1 and 2.[2]

Table 1. Period distribution in the Luganda corpus

Period	Tokens	%
1890s	39,538	0.98
1900s	310,548	7.66
1910s	228,198	5.63
1920s	144,776	3.57
1930s	293,433	7.24
1940s	120,395	2.97
1950s	413,398	10.20
1960s	219,428	5.41
1970s	167,377	4.13
1980s	243,978	6.02
1990s	168,746	4.16
2000s	724,317	17.87
2010s	979,607	24.17
TOTAL	**4,053,739**	**100.00**

Figure 1 combines both topic/genre and period distributions. It shows how individual topics/genres included in the corpus (x-axis) are represented, in actual number of tokens (y-axis), across the different time periods (z-axis). Furthermore, there is a relatively even distribution of topics/genres throughout the different time periods. For example, apart from two time slots (i.e. the 1890s and 1910s) in which only three topics/genres are included, all the remaining periods include at least four topics/genres. The most current periods (i.e. the 2000s and 2010s), however, include the highest number of topics/genres; namely, 13 and 16 respectively, out of the total of 18 topics/genres. This is, understandably, due to recent advances in the writing of the Luganda language on a wide range of topics.

In general, corpora for Bantu languages are "raw" corpora, in the sense that the orthographic words in the texts are not accompanied by their grammatical class: Bantu corpora have not been POS-tagged (the tokens have not been tagged for their parts of speech), nor have they been tagged for any other features. The only known exceptions for Bantu are two POS-tagged corpora for Swahili and

2. The Luganda corpus built for the purpose of the present study may be consulted at *BantUGent – UGent Centre for Bantu Studies*, at Ghent University. For more information, please see https://research.flw.ugent.be/en/bantugent.

Table 2. Topic/genre distribution in the Luganda corpus

Topic/Genre	Tokens	%
Agricultural Documents	111,928	2.76
Cultural Texts	287,390	7.09
Environmental Documents	37,229	0.92
Financial Texts	78,742	1.94
Folktales	207,333	5.10
Health Documents	92,756	2.29
Historical Texts	617,042	15.22
Inspirational Materials	15,234	0.38
Instructional Materials	450,582	11.12
Legal Texts	183,596	4.53
Magazines	143,099	3.53
Newspapers	615,341	15.18
Novels	373,410	9.21
Plays	76,746	1.89
Political Documents	746,106	1.14
Radio News	7,321	0.18
Religious Texts	704,675	17.38
Songs	5,981	0.15
TOTAL	**4,053,739**	**100.00**

three smaller POS-tagged corpora for Northern Sotho, Zulu and Cilubà (De Pauw et al., 2012). The Luganda corpus is thus a raw corpus. To query the Luganda corpus, version 7 of the *WordSmith Tools* software suite was used (Scott, 2019) and the search results (concordance lines) were then exported to an *Excel* spreadsheet for further analysis. For each of the modal markers under question, its basic form was searched for, surrounded by stars to cater for the possible inflections. (To illustrate the process, an example for one of the modal markers is given at the start of Section 4.) For each of the modal markers, the full corpus was first queried. Based on the number of results, a new search was launched using the random "Sample" function of *WordSmith*, with the sample rate chosen in such a way that "enough" concordance lines could be analysed, typically up to about one thousand lines. When occurrence frequencies are mentioned below, these are the extrapolations from the samples.[3]

3. Given the large number of periods (13 decades in all), it is the case that some periods end up with relatively small samples. However, we did not do stratified sampling to ensure a minimum

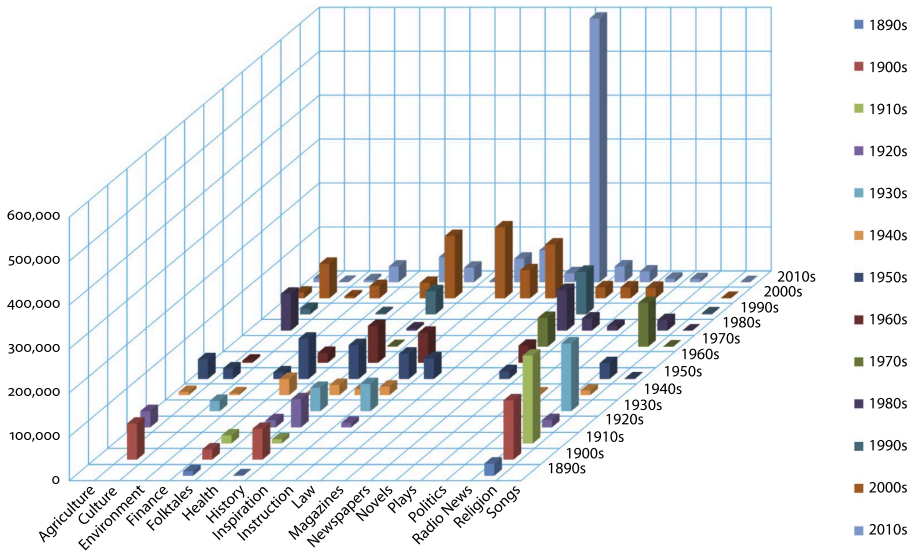

Figure 1. Combined topic/genre and period distribution in the Luganda corpus

Also note that in Luganda tones are not marked in the orthography, although tone is distinctive both lexically and grammatically. For this reason, tone marks were added to the examples used in this article. This was only systematically done on the surface level, i.e. the first example line. Since underlying tone and tone rules are not yet well established for Luganda, tone is not marked in the morphological parsing line. Additionally, the lexical tone of the modal markers cited in the running text of this article were also marked.

4. An account of *-yînz-*

In this section, we illustrate the different uses of *-yînz-* with examples drawn from different sections of the corpus. It will be shown that *-yînz-* expresses modal meanings that indicate whether the state of affairs expressed by the main verb of the utterance is possible, and that *-yînz-* expresses lexical meanings that refer to an actual state of affairs in the outside world.

To examine the distribution and uses of *-yînz-*, the corpus was queried with the following search items: **yinz*/*ainz*/*einz*/*iinz*/*oinz*/*uinz** in order to

number of lines per period, as similar work on Kirundi (Mberamihigo, 2014) taught us that the analysis of samples of several thousands of lines basically gave us the same results as the analysis of a sample of about 1,000 lines.

cater for all the possible spelling conventions, both current and historical, and prefix concord conjugations. To reduce the number of undesired search results, such as nouns or unrelated verbs, the items: **buyinza/*buinza/*inzika/omuyinza/omuinza*, most of which are deverbative nouns, were blocked using the "Advanced" function in *WordSmith's* "Concord" tool. First, we note that *-yînz-* occurs 7,462 times in the 4-million-word corpus. It thus occurs on average 18.4 times for every 10,000 words. We analysed a sample of these, totalling 840 instances, and the analysis of these formed the basis for the current distributional description. Overall, as can be seen in Figure 2, *-yînz-* is almost exclusively used in double-verb constructions, that is, where it is followed by a second verb in the infinitive. It is used in such constructions in 97.5% of the cases. In only 0.3% of the cases is it used outside double-verb constructions ("Lexical" in Figure 2), that is, where it is used both without a second verb and where it clearly expresses its original lexical meanings. In 2.2% of the cases the infinitival main verb is moved from its canonical position or deleted altogether but remains simply implied ("Lexical^{+}" in Figure 2). Double-verb constructions will, in the following description, be considered as instances where *-yînz-* carries modal meanings, while cases labelled "Lexical" and "Lexical^{+}" represent, respectively, those instances in which *-yînz-* clearly expresses its original lexical meanings and those in which it ambiguously carries lexical and modal meanings.

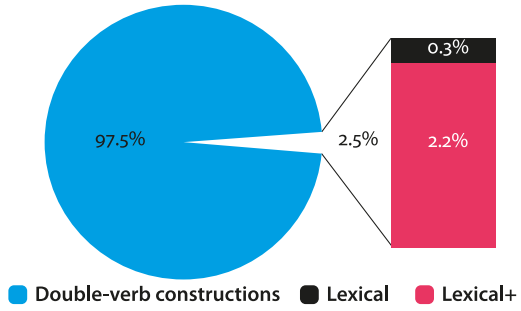

Figure 2. Environments in which *-yînz-* is found

4.1 The modal uses of *-yînz-*

In double-verb constructions, *-yînz-* is mostly involved in the expression of epistemic possibility, expressing this meaning 2,692 times (37%). In second place is deontic possibility, the expression of moral desirability or permission (see Nuyts 2006:5), expressed 1,897 times (26%). This is followed by situational dynamic possibility and participant-inherent dynamic possibility, expressed, respectively, 1,307 times (18%) and 868 times (12%). When expressing participant-imposed

dynamic possibility, *-yînz-* occurs 508 times (7%). Overall, deontic and epistemic possibility together constitute a bigger part (63%) of the modal domain of *-yînz-* than the dynamic possibility uses (participant-inherent, participant-imposed and situational dynamic possibility, which constitute 37%). This distribution is summarised in Figure 3 below.

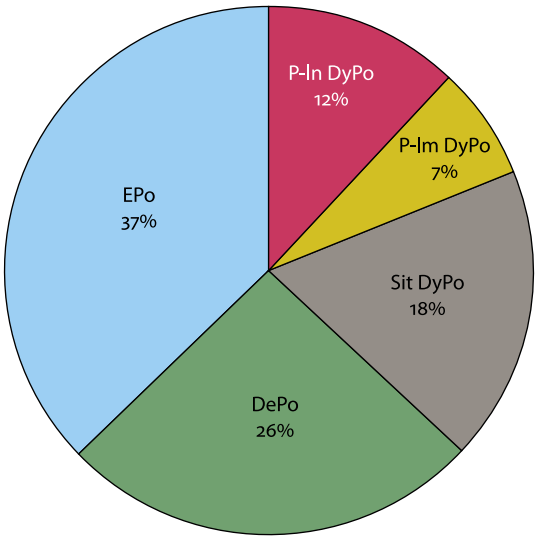

Figure 3. Modal uses of *-yînz-*, as seen in the entire corpus

When expressing participant-inherent dynamic possibility, as in (1) to (3), *-yînz-* is followed by a verb in the infinitive. In (1), the infinitival main verb *okwógérá* is followed by an object noun phrase, while in (2) and (3) an object prefix is attached to the infinitival main verb. In (1), the speaker uses *-yînz-* to express an inherent ability of an Indian to speak English, which often earns them respect among the people of their country, while in (2), the speaker uses *-yînz-* to show that he has the capacity to give to the listener medicine that can make another person crazy and become completely senseless. In (3), it is shown that only the subject referent of the clause, and no one else, has the ability to remove them from the pillory. From the context, it is the metaphorical pillory (i.e. sin) that only (Catholic) priests are inherently able to take others out of, by virtue of their being successors of Peter.

(1) Omuîndi atérá ókúsíbwámú ekitîbwa abéwabwe, bwabá nga aînzá ókwógérá Ólúngéréza.
o-mu-yindi a-Ø-ter-a
AUG_1-NP_1-Indian SP_1-PRS-be_often-IPFV
o-ku-ss-ibw-a-mu e-ki-tiibwa a-ba-a
AUG_{15}-NP_{15}-put-PASS-FV-LOC_{18} AUG_7-NP_7-respect AUG_2-PP_2-CONN
e-wa-abwe bwe a-ba nga **a-Ø-yinz-a**
AUG_{23}-PP_{23}-$POSS_2$ if SP_1-be that **SP_1-PRS-POT-IPFV**
o-ku-oger-a o-lu-ngereza
AUG_{15}-NP_{15}-speak-FV AUG_{11}-NP_{11}-English
"An Indian is often respected by the people of his place if he can speak English."

(Olusuku, Folktales, 1910s)[4]

(2) [...] até wáliwó n'ékirala, nnyînzá ókúkúwa eddágálá né túmúsíríwázá ne kífúúkira ddálá kísíru [...]
ate wa-Ø-li-wo ne e-ki-lala **N-Ø-yinz-a**
and SP_{16}-PRS-be-LOC_{16} also AUG_7-PP_7-other **SP_{1SG}-PRS-POT-IPFV**
o-ku-ku-w-a e-ddagala ne
AUG_{15}-NP_{15}-OP_{2SG}-give-FV AUG_5-NP_5.medicine and
tu-Ø-mu-siriwal-y-a ne
SP_{1PL}-PRS-OP_1-become_crazy-CAUS-IPFV and
ki-Ø-fuuk-ir-a ddala ki-siru
SP_7-PRS-become-APPL-IPFV completely PP_7-senseless
"[...] and there is also another thing, I can give you medicine and we make him crazy and he becomes completely senseless [...]"

(Pawulo, Novels, 1980s)

(3) Fé feká túyínzá okubágyá mú nvúbá omwó.
ffe ffeka **tu-Ø-yinz-a** **o-ku-ba-ggy-a** mu
we we_only **SP_{1PL}-PRS-POT-IPFV** **AUG_{15}-NP_{15}-OP_{2PL}-take_out-FV** LOC_{18}
N-vuba o-mu-o
NP_9-pillory AUG_{18}-PP_{18}-DEMb
"Only we can take you out of that pillory."

(Anoonya, Religious Texts, 1890s)

In Examples (4) and (5), *-yînz-* expresses participant-imposed dynamic possibility. In (4), where the main verb is immediately followed by a prepositional phrase, the ability to reach the shore safely is made possible by the shallowness of the

4. For an explanation of the abbreviations and symbols used in the interlinear glossing, see Appendix A.

water in which they fell, but it is not due to their own ability or capacity. Similarly in (5), where the main verb is followed by an object noun phrase, it is because the country is fertile that the people (would have) had the ability to grow food. The people alone, without the enabling factor of fertility, are not able to grow food. In the same example, *-yînz-*, in the perfective, is prefixed with the conditional affix *-andi-* in what appears to be a counterfactual conditional construction with an elided protasis (see Kawalya et al., 2018a). The assumed elided protasis also provides a condition which, if fulfilled, would be an enabling factor determining the ability of the subject referent to grow food. Thus, in Example (5), there are two kinds of external factors triggering a participant-imposed dynamic possibility reading, viz. the fertility of the country and the fulfilment of an implicit condition in the elided protasis.

(4) Tewááli n'ómú yafâ erá nga túyinzá ókútúuká kú lúkálu n'éddémbé; ánti amázzi mwe twáyiika gáálí gá ffúúti ssátu.
te-wa-a-li ne o-mu a-a-f-a era nga
NEG-PP_{16}-REM_PST-be with PP_1-one PP_1-REM_PST-die-PFV and then
tu-yinz-a **o-ku-tuuk-a** ku lu-kalu ne
SP_{1PL}-POT-IPFV **AUG_{15}-NP_{15}-reach-FV** LOC_{17} NP_{11}-shore with
e-ddembe anti a-ma-zzi mu-e tu-a-yiik-a
AUG_5-NP_5.peace as AUG_6-NP_6-water PP_{18}-REL SP_{1PL}-REM_PST-fall-PFV
ga-a-li ga-a ffuuti ssatu
SP_6-REM_PST-be PP_6-CONN NP_{10}.foot three
"There was no one who died and then we could reach the shore safely, as the water in which we fell was three feet deep."

(Ekizinga, Folktales, 1950s)

(5) Ensí eno ngimû; embéérá y'óbúdde nnúngi, twándíyínzízzá okulímá émmére ne túgíkúngula erá ne túgíryâ […]
e-N-si e-no N-gimu e-N-beera ya-a
AUG_9-NP_9-country PP_9-DEMa PP_9-fertile AUG_9-NP_9-state PP_9-CONN
o-bu-dde N-lungi **tu-andi-yinz-ye** **o-ku-lim-a**
AUG_{14}-NP_{14}-time PP_9-good **SP_{1PL}-IRR-POT-PFV** **AUG_{15}-NP_{15}-grow-FV**
e-mmere ne tu-Ø-gi-kungul-a era ne
AUG_9-NP_9.food and SP_{1PL}-PRS-OP_9-harvest-IPFV even and
tu-gi-ly-a
SP_{1PL}-OP_9-eat-IPFV
"This country is fertile; the weather is good, we would have been able to grow food and even harvest it and eat it […]"

(Amaka, Novels, 1980s)

Example (6) shows *-yînz-* expressing situational dynamic possibility. In this sentence, where the main verb is followed by a prepositional phrase, *-yînz-* does not show an ability or capacity of the first-argument participant in the clause, in this case the inanimate subject *emméerí* "ships". It rather signifies a general possibility inherent in the situation as a whole, that is, that the situation at hand makes it possible for big ships to enter into the inside of the harbour.

(6) Gwe mwâlo mulúngi, erá emêri enéne ezíita ku liyánja kakáno zíinzá kúîngírá mú mwâló ogwo mundá.
gu-e mu-alo mu-lungi era e-mmeeri e-N-nene
PP_3-REL NP_3-harbour PP_3-good and AUG_{10}-NP_{10}.ship AUG_{10}-PP_{10}-big
e-zi-yit-a ku li-yanja kaakano **zi-Ø-yinz-a**
AUG_{10}-PP_{10}-pass-FV LOC_{17} NP_5-lake now **SP_{10}-PRS-POT-IPFV**
ku-yingir-a mu mu-alo o-gu-o mu-nda
NP_{15}-enter-FV LOC_{18} NP_3-harbour AUG_3-PP_3-DEMb PP_{18}-inside
"It is a good harbour, and big ships that move on the sea can now enter into the inside of that harbour."
(Enfaanana, Historical Texts, 1920s)

In (7) and (8), *-yînz-* expresses deontic possibility. In (7), it is still followed by the infinitive of the main verb which is in turn followed by an object noun phrase, while in (8) it is followed by the infinitive of the auxiliary *-ba* "to be" which functions as the main verb and is immediately followed by an object complement. In (7), according to societal naming practices, it is acceptable to give certain names to both boys and girls; and in (8), the writer points out that it is acceptable to appoint a doctor or a teacher or an agricultural advisor or a veterinary doctor.

(7) Amánnyá ágo gáyinzá ókútúúmwá ábálénzí n'ábáwála […]
a-ma-nnya a-ga-o **ga-Ø-yinz-a**
AUG_6-NP_6-name AUG_6-PP_6-DEMb **SP_6-PRS-POT-IPFV**
o-ku-tuum-w-a a-ba-lenzi ne a-ba-wala
AUG_{15}-NP_{15}-name-PASS-FV AUG_2-NP_2-boy and AUG_2-NP_2-girl
"Those names may be given to boys and girls […]"
(Amannya, Cultural Texts, 1980s)

(8) Omúntú álóndébwá ayînzá ókúbá Omusáwo, obó Omusómésa, obá Omuwí w'ámágézi ag'éby'ókulima, obá Omusáwó w'ébisóló […]
o-mu-ntu a-Ø-lond-ebw-a **a-Ø-yinz-a**
AUG_1-NP_1-person SP_1-PRS-appoint-PASS-IPFV **SP_1-PRS-POT-IPFV**
o-ku-ba o-mu-sawo oba o-mu-somesa oba
AUG_{15}-NP_{15}-be AUG_1-NP_1-doctor or AUG_1-NP_1-teacher or
o-mu-wi wa-a a-ma-gezi a-ga-a
AUG_1-NP_1-giver PP_1-CONN AUG_6-NP_6-knowledge AUG_6-PP_6-CONN
e-bi-a o-ku-lim-a oba o-mu-sawo wa-a
AUG_8-PP_8-CONN AUG_{15}-NP_{15}-farm-FV or AUG_1-NP_1-doctor PP_1-CONN
e-bi-solo
AUG_8-NP_8-animal
"A person who is appointed may be a doctor or a teacher or an agricultural advisor, or a veterinary doctor […]"

(Gavumenti, Political Documents, 1940s)

Lastly, in (9) *-yînz-* expresses epistemic possibility. The speaker, in this sentence, expresses a likelihood that the people he is going to see will fear and not say anything to them if they went together. Note that this construction is different from the one in Kirundi where the modal verb *-shóbor-* is involved in the expression of epistemic possibility. In Kirundi, for *-shóbor-* to express epistemic possibility, it always has to combine with the infinitive of the auxiliary *-ba* "to be" which is in turn followed by the main verb (Bostoen et al., 2012: 17). Luganda has a similar construction, as shown in (10), where *-yînz-* is followed by the infinitive of the auxiliary *-ba* "to be". However, in Luganda this is not immediately followed by a finite main verb. In this example, it is followed by a complement clause introduced by *nga* "that".

(9) Bwetúgendá ffembî bé nngénda okulába báyinzá ókútyá ne bákóná, ne bátámbúúlira kintú kyonnâ.
bwe tu-Ø-gend-a ffembi ba-e N-gend-a
if SP_{1PL}-PRS-go-IPFV we_both PP_2-REL SP_{1SG}-go-IPFV
o-ku-lab-a **ba-Ø-yinz-a** **o-ku-ty-a** ne
AUG_{15}-NP_{15}-see-FV **SP_2-PRS-POT-IPFV AUG_{15}-NP_{15}-fear-FV** and
ba-Ø-kon-a ne ba-ta-N-buulir-a ki-ntu ki-onna
SP_2-PRS-keep_silent-IPFV and SP_2-NEG-OP_{1SG}-tell-IPFV NP_7-thing PP_7-any
"If we go together, the people whom I am going to see may fear and keep quiet, and not tell me anything."

(Guluma, Novels, 1990s)

(10) Ebigámbó ebyo Katíkkiro, ayînzá ókúbá nga yabíwúlírá búlúngi.
e-bi-gambo e-bi-o Ø-katikkiro
AUG_8-NP_8-word AUG_8-PP_8-DEMb NP_1-Prime_Minister
a-Ø-yinz-a **o-ku-ba** nga a-a-bi-wulir-a
SP_1-PRS-POT-IPFV AUG_{15}-NP_{15}-be that SP_1-REM_PST-OP_8-hear-PFV
bu-lungi
PP_8-well
"Those words, the Prime Minister, he might have heard them well."
(Amazina, Historical Texts, 1990s)

In all the examples above, *-yînz-* is always followed by a second verb in the infinitive. This second verb functions as the main verb, which refers to the main event of the clause.

4.2 The lexical uses of *-yînz-*

In our corpus sample, we found only two examples where *-yînz-* carries its original lexical meanings as presented in older literature on Luganda. In (11), *-yînz-*, which is followed by an emphatic adverb *nnyo* "very", can be translated as "be powerful"; while in (12), where it is followed by the noun phrase *buli kantu* "everything", it is translatable as "overcome, manage, control".

(11) Okusábá kw'ómuntú ómútúkirivu kúyínzá nyó mu kukólá kwákwô.
o-ku-sab-a ku-a o-mu-ntu o-mu-tuukirivu
AUG_{15}-NP_{15}-pray-FV PP_{15}-CONN AUG_1-NP_1-person AUG_1-NP_1-righteous
ku-Ø-yinz-a **nnyo** mu ku-kol-a ku-akwo
SP_{15}-PRS-be_powerful-IPFV very LOC_{18} NP_{15}-work-FV PP_{15}-$POSS_{15}$
"The prayer of a righteous person is very powerful in the way it works."
(Katonda, Religious Texts, 1950s)

(12) Néwákúbáddé abantu básásúlá gávúmênti emisólo okubáyambá kú bábákóla obubî, kiremé okulówóózébwá nti gávúmênti eyînzá buli kantu […]
newakubadde a-ba-ntu ba-Ø-sasul-a Ø-gavumenti
although AUG_2-NP_2-person SP_2-PRS-pay-IPFV NP_9-government
e-mi-solo o-ku-ba-yamb-a ku ba-Ø-ba-kol-a
AUG_4-NP_4-tax AUG_{15}-NP_{15}-OP_2-help-FV LOC_{17} SP_2-PRS-OP_2-do-IPFV
o-bu-bi ki-rem-e o-ku-lowooz-ebw-a nti
AUG_{14}-NP_{14}-bad SP_7-not_be-SBJV AUG_{15}-NP_{15}-assume-PASS-FV that
Ø-gavumenti **e-Ø-yinz-a** **buli ka-ntu**
NP_9-government **SP_9-PRS-control-IPFV every NP_{12}-thing**
"Although people pay taxes to government to help them with those who do them bad, it should not be assumed that government controls everything […]"
(Okukula, Inspirational Materials, 1940s)

Although *-yînz-* is generally overlooked in Luganda dictionaries and grammars, unlike its other modal counterparts *-sóból-* and *-andi-*, one dictionary (Le Veux, 1917) provides a relatively comprehensive account of it. First of all, Le Veux considers *-yînz-* to be a causative form of the verb *-yîng-* which he translates as *excéder* "exceed" as in Example (13), or *dépasser* "surpass" as in (14). In Luganda, the short causative extension *-y-* causes spirantisation to preceding segments ending in "d, t, k or g" (Ashton et al., 1954: 153; Bostoen, 2008: 322). Presently, *-yînz-* should no longer be considered a causative form of *-yîng-*, but rather a lexicalised causative form of the verb, as it has developed its own particular meanings in which the derivational causative meaning is no longer fully noticeable with regard to the original base verb. Ashton et al. (1954: 998) and Bostoen (2008: 335) quoting Bastin et al. (1983), further show that Luganda manifests spirantisation before the perfect(ive) suffix (*-ye*) in the same way it does before the causative suffix (*-y(a)*). Indeed, as shown in (13) and (14), the perfective of *-yîng-* (i.e. *-yinz-e*) looks like its causative form.

(13) Eddalú lye líyinzé.
e-ddalu li-e **li-Ø-ying-ye**
AUG_5-NP_5.stubbornness PP_5-$POSS_1$ **SP_5-PRS-exceed-PFV**
"His stubbornness is excessive."
(Le Veux, 1917: 997)

(14) Bátúyînze obúngi.
ba-Ø-tu-ying-ye o-bu-ngi
SP_2-PRS-OP_{1PL}-surpass-PFV AUG_{14}-NP_{14}-quantity
"They have surpassed us in quantity (number)."
(Le Veux, 1917: 997)

Le Veux (1917: 998) translates *-yînz-* as *violenter* "to abuse, assault (sexually)", *être plus fort, plus nombreux que* "to be stronger, more numerous than", *venir à bout de, pouvoir* "to manage to overcome/finish, can" and *être qualifié pour* "to be qualified for", providing Examples (15) to (17). The nominal forms *obuyinza* (*pouvoir* "power") and *omuyinza* (*capable, puissant* "capable, powerful") are also derived from *-yînz-*.

(15) Omusájja yánnyinzá.
o-mu-sajja **a-a-N-yinz-a**
AUG_1-NP_1-man **SP_1-REM_PST-OP_{1SG}-abuse-PFV**
"The man abused me."
(Le Veux, 1917: 998)

(16) Tibájjá kutúyinzá.
ti-ba-Ø-jj-a **ku-tu-yinz-a**
NEG-SP_2-PRS-come-IPFV **NP_{15}-OP_{1PL}-overcome-FV**
"They are not going to defeat us."

(Le Veux, 1917: 998)

(17) Bátúyínzízzâ.
ba-Ø-tu-yinz-ye
SP_2-PRS-OP_{1PL}-defeat-PFV
"They have defeated us."

(Le Veux, 1917: 998)

In (15) to (17), *-yînz-* is used, just like in (11) and (12), outside a double-verb construction. It appears autonomously as the main verb expressing the main event in the clause. Moreover, in Examples (15) to (17), an object prefix is attached to *-yînz-* itself and not to another verb in the clause, as is the case in several examples in the previous Section 4.1. Its capacity to incorporate an object prefix indicates that it is a transitive verb, which is in line with its causative etymology. In addition, Le Veux provides the only examples we have seen where *-yînz-* is immediately followed by a direct object. In both (18) and (19) it can be translated as "manage (to finish / to carry)".

(18) Síjjá kuyînzá múlímó ogwó.
si-Ø-jj-a **ku-yinz-a** **mu-limo** o-gu-o
NEG.SP_{1SG}-PRS-come-IPFV **NP_{15}-manage-FV** **NP_3-work** AUG_3-PP_3-DEMb
"I will not manage (to finish) that work."

(Le Veux, 1917: 732)

(19) Onóóyînzá ómúgúgu guno?
o-noo-yinz-a **o-mu-gugu** gu-no
SP_{2SG}-NEAR_FUT-manage-IPFV **AUG_3-NP_3-burden** PP_3-DEMa
"Will you manage (to carry) that burden?"

(Le Veux, 1917: 806)

4.3 Diachronic distributional analysis of *-yînz-*

We considered all 13 decades at our disposal (1890s to 2010s) to study the diachronic semantic distribution of *-yînz-*. While the occurrence frequency of *-yînz-* oscillates wildly with time, there are three phases in the data: between 10–15 per 10,000 words until the 1910s, then a sharp increase to a second phase from the 1920s to 1950s, and then a decline to around 15 again from the 1960s to the present. There has thus been relative stability since the 1960s, for more than 50 years. See Figure 4.

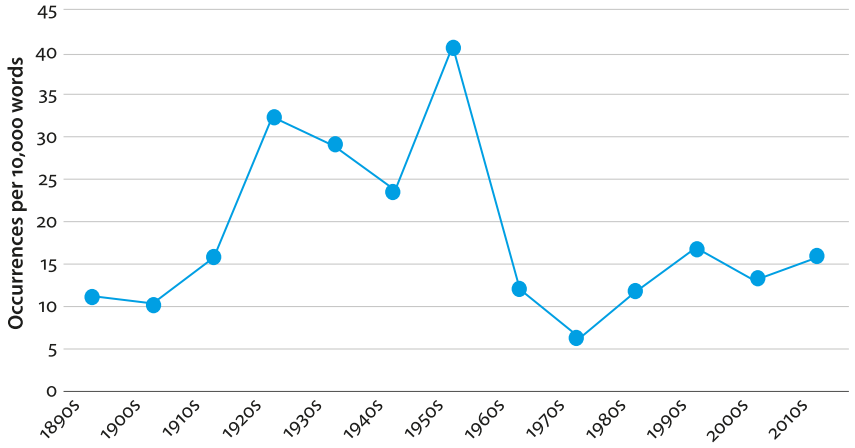

Figure 4. Overall occurrence frequency of *-yînz-* through time

One decade in particular, the 1950s, stands out, given that *-yînz-* occurs about 40 times for each 10,000 words in that period. In an attempt to pinpoint the cause of this outstanding and rather irregular frequency of *-yînz-*, we looked at the internal composition of the corpus material in the 1950s. It was found that although material for this decade was drawn from 22 different texts, over one third (35%) of all the occurrences of *-yînz-* in this decade are from a single text. Furthermore, nine out of ten (91%) of all instances of *-yînz-* in this text express deontic possibility. This text, which is also responsible for two thirds (65%) of the total number of instances of deontic possibility in the 1950s, is a translated legal book, wherein *-yînz-* was generally used to translate deontic *may* which appears very frequently in the original English version. This clearly implies that the frequency for *-yînz-* in general and deontic possibility in particular for the 1950s is an overrepresentation.

A detailed picture of the distribution of the various meanings of *-yînz-* over time is shown in Figures 5 and 6. Figure 5 depicts the actual occurrences per 10,000 words, while Figure 6 sets out each decade in percentage.

As was already clear from Figure 2, as well as from the discussion presented in Section 4.2, *-yînz-* is almost not found as a main verb expressing lexical meanings. The very few instances that are attested in our sample are from the 1940s and 1950s. From the 1890s up to the 1940s, *-yînz-* also carries meanings which are ambiguous between lexical and participant-inherent dynamic possibility (i.e. Lexical^{+}). Another clear manifestation is that participant-inherent dynamic possibility meanings used to be prominent in the earlier decades, but their share (see Figure 6) continued to reduce over time and are almost disappearing from the semantic range of *-yînz-*. For participant-imposed dynamic possibility meanings

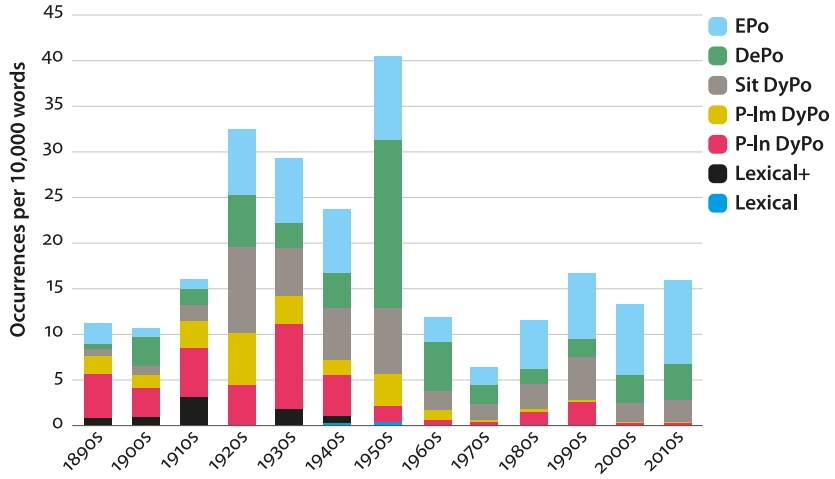

Figure 5. Distribution of the actual occurrences of the uses of *-yînz-* through time[5]

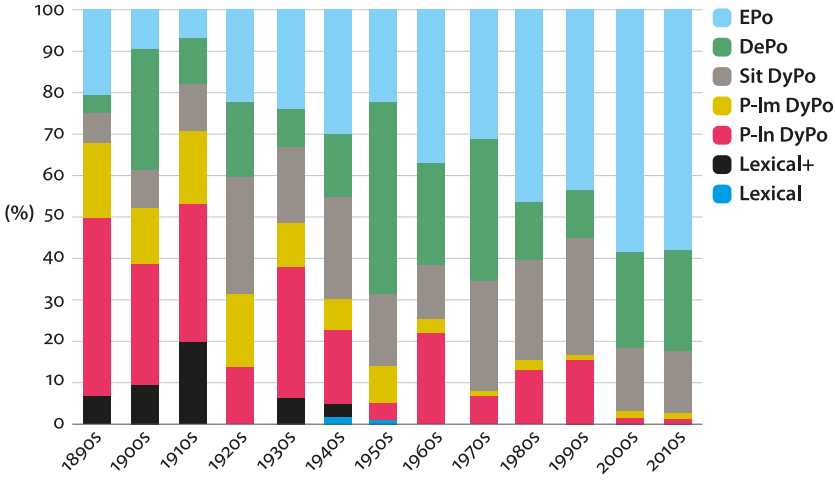

Figure 6. Percentage distribution of the uses of *-yînz-* through time

5. The raw frequency facts for this figure, as well as all comparable ones following, are to be found as online supplementary material (see Appendix B). Two-tailed Fisher's exact tests reveal that most comparisons between adjacent columns (so between neighbouring decades) are not statistically relevant, but such tests for columns (and thus decades) further apart are statistically relevant.

one notices a break: their occurrence in the meaning distribution is stable from the 1890s to the 1950s, after which they are not meaningfully represented anymore as they are attested less than once for every 10,000 words between the 1960s and 2010s. Situational dynamic possibility and deontic possibility are relatively stable meanings through time. This should not be surprising given that the two categories share some common features according to van der Auwera & Plungian (1998), who subsume Nuyts' situational dynamic modality under what they call participant-external modality and view deontic modality as "a sub-domain or special case of participant-external modality" (van der Auwera & Plungian, 1998: 81). The two categories are concerned with enabling or compelling circumstances external to the participant. However, if one considers the three sub-categories of dynamic possibility together, it is evident that as a whole the importance of dynamic possibility in the semantic range of *-yînz-* has seriously shrunk over the last 130 years, i.e. from 75% in the 1890s to only 17% today. Conversely, epistemic possibility meanings are seen to grow continuously over the years: from about 10% of all cases a century ago, to about 60% today. In other words, the semantic core of the modal marker *-yînz-* has undergone strong 'subjectification' in that it has unmistakably shifted from 'objective' meanings primarily referring to the outside world to 'subjective' meanings based in the speaker's belief towards the externally described situation (see Traugott, 1989: 35).

4.4 Is *-yînz-* an auxiliary or not?

Anderson (2006: 5) defines an auxiliary as "an element that in combination with a lexical verb forms a monoclausal verb phrase with some degree of (lexical) semantic bleaching that performs some more or less definable grammatical function". In what precedes, we have seen that *-yînz-* indeed mostly constitutes a monoclausal verb phrase in combination with a main verb in the infinitive. It is rarely found in constructions where it is not followed by another verb. In all those true double-verb constructions, *-yînz-* is restricted to expressing meanings within the possibility subdomain of modality. The second verb, and not *-yînz-*, represents the main event in the clause; *-yînz-* only shows that that event is possible. In this case, it is clear, following Anderson, that *-yînz-* functions as an auxiliary carrying modal meanings.

Moreover, in almost all Luganda references we consulted, *-yînz-* is defined as a modal verb with the meanings "be able (to handle, to do), be capable" (see Mulira & Ndawula, 1952: 118; Chesswas, 1963: 89; Cole, 1967: 119, 128; Snoxall, 1967: 348; Murphy, 1972: 630). Only one dictionary (Le Veux, 1917) defines and uses it as a lexical verb, as was seen in Examples (15) to (19). This focus on lexical uses is not unexpected in a century-old lexicographical resource. Later dictionaries do

report it as a modal (see Mulira & Ndawula, 1952: 118; Snoxall, 1967: 348; Murphy, 1972: 630).

However, like *-sóból-* and similar auxiliaries in Luganda and many other languages, *-yînz-* shares a number of formal characteristics with the rest of the verbs in the language. As seen in the different examples above and just as its near-synonym *-sóból-* (Kawalya et al., 2014: 79), *-yînz-* can be inflected for tense and aspect; it can occur in non-finite forms like the infinitive; it can form nouns or be used as a noun itself; it can take a direct object or object concords; and it can also be modified by intensifiers.

That notwithstanding *-yînz-*, just like *-sóból-*, can be differentiated from the other verbs in one important respect. When expressing modality *-yînz-* is almost always found in double-verb constructions, with a main verb in the infinitive. But, unlike *-sóból-* whose main verb can be omitted and only referred to by an object prefix on the auxiliary when expressing dynamic possibility (Kawalya et al., 2014: 79), this is not possible with *-yînz-*. In fact, for the only examples in which *-yînz-* was found to carry an object prefix, in (15) to (17), *-yînz-* carries lexical meanings and the object prefixes are actually object personal pronouns.

However, we did find constructions with meanings that are ambiguous between lexical and participant-inherent dynamic possibility, where the main verb is omitted and thus only implied. In (20) and (21), *-yînz-*, which occurs sentence-finally, appears without another verb. In (20), it appears as if the main verb is *kulamula* "to rule" which is mentioned in the preceding sentence (*Nkuze sikyainza kulamula Buganda* "I have grown old; I can no longer manage to rule Buganda"), and that the speaker just avoids repeating it in the following sentence. The dynamic possibility meaning of this sentence is obtained when one assumes, from the context, that *kulamula* is the main verb. On the other hand, if we do not appeal to context, *-yînz-* is also interpretable as the main verb and can be translated as "be qualified". In (21), the infinitive *ókúgábírá* "to give to" could be considered as the main verb of the auxiliary *-yînz-*. Since it was previously mentioned, it is not repeated and left implicit. On the other hand, here again, the use of *-yînz-* is close to that of a main verb with the lexical meaning "be qualified for".

(20) N'ágámba nti; abána bange, obugânda buno mubulyê, nze síkyáînzá.
ne a-gamb-a nti a-ba-ana ba-ange o-bu-ganda
and SP_1-say-IPFV QUOT AUG_2-NP_2-child PP_2-$POSS_{1SG}$ AUG_{14}-NP_{14}-Ganda
bu-no mu-bu-ly-e nze **si-kya-yinz-a**
PP_{14}-DEMa SP_{2PL}-OP_{14}-take_over-SBJV me **NEG.SP_{1SG}-PERS-POT-IPFV**
"And he said: my children, this Buganda, you should take it over, I can no longer manage (to rule it)."

(Engero3, Folktales, 1890s)

(21) Erá nayé buli muntu kímugwánírá ókúgábírá énýó abávu ku bintú byálínâ ngá bwáyínzâ.
era naye buli mu-ntu ki-Ø-mu-gwan-ir-a
and but every NP_1-person SP_7-PRS-OP_1-require-APPL-IPFV
o-ku-gab-ir-a ennyo a-ba-avu ku bi-ntu
AUG_{15}-NP_{15}-give-APPL-FV much AUG_2-NP_2-poor LOC_{17} NP_8-thing
bi-e a-Ø-lin-a nga bwe **a-Ø-yinz-a**
PP_8-REL SP_1-PRS-have-IPFV as how **SP_1-PRS-POT-IPFV**
"But for every person it is required to give to the poor as much from the things he has as he can."
(Ekitabo eky'okusaba, Religious Texts, 1910s)

In (20) and (21), therefore, *-yînz-* appears to be structurally outside a double-verb construction, but from the context a main verb seems to just have been omitted. What is more important, however, is that these sentences convey meanings that border on lexical and participant-inherent dynamic possibility (hence Lexical$^+$) – meanings that are conceptually too close to each other to tease them apart. On the other hand, when clearly expressing modality, *-yînz-* is always used together with a main verb in the infinitive; the main verb can never be represented by an object concord on the auxiliary or completely omitted. This inability of *-yînz-* to (clearly) convey modality when used alone, without an accompanying main verb, shows that when expressing modality, it loses its morphosyntactic independence and behaves more like a true auxiliary required to always co-occur with a second verb. It also suggests that increased grammaticalisation in the syntactic domain goes hand in hand with subjectification in the semantic domain. The second verb can only be omitted when *-yînz-* is expressing lexical meanings or when its meaning is ambiguous between lexical and the more objective dynamic possibility meanings – meanings based in the externally-described situation (Traugott, 1989). However, it is obligatorily present in the case of the clear instances of dynamic possibility and the more subjective deontic and epistemic possibility meanings – meanings which "tend to become based in the speaker's subjective belief state / attitude toward the proposition" (Traugott, 1989: 35).

5. An account of *-sóból-*

Kawalya et al. (2014) show that the modal verbs *-sóból-* and *-yînz-* are the two most frequent markers of possibility, others being *-andi-*, *-sóbózés-*, *-sóbók-* and *-sóbós-*. They furthermore demonstrate that the use of *-sóból-* as an auxiliary in double-verb constructions used to be low in earlier periods but grew over time

and became more strongly associated with the expression of modality in more recent times. What is more, the semantic range of *-sóból-* was extended from dynamic possibility to deontic possibility from the 1960s onwards. Given that the study of the possibility modal verb *-sóból-* in Kawalya et al. (2014) was based on a corpus of just 1.5 million words, the former distributional analyses have been repeated on the basis of the current 4-million-word corpus. A summary of the new findings is presented below.[6]

5.1 Distributional analysis of *-sóból-*

According to both Kawalya et al. (2014) and our current analysis, *-sóból-* is involved in the expression of lexical meanings as well as modal meanings. The modal meanings of *-sóból-* are participant-inherent dynamic possibility, participant-imposed dynamic possibility, situational dynamic possibility and deontic possibility. As shown in Figure 7, lexical meanings contribute 4.5% to the entire semantic range of *-sóból-*. Participant-inherent dynamic possibility meanings contribute 17.5%, while participant-imposed dynamic possibility contributes 26%. Situational dynamic possibility contributes 41% and deontic possibility contributes only 11%.

5.2 Diachronic distributional analysis of *-sóból-*

Overall, *-sóból-* occurs 6,782 times in the 4-million-word corpus, or 16.7 times for every 10,000 words. This average is of little value, however, as there has been a major increase in its growth in the most recent decades compared to about a century ago, as seen in Figure 8. It occurs at a very low frequency of about once for every 10,000 words in the earlier decades (1890s to 1920s) and attains a slow growth to stabilise at about 10 times for every 10,000 words from the 1950s to 1990s, before a major increase to about 30 times for every 10,000 words in the most recent decades (2000s and 2010s).

Figures 9 and 10 show the actual meaning distributions of *-sóból-* for each decade anew – actual occurrences per 10,000 words in Figure 9, and set out in percentage in Figure 10. Given the growth in the use of *-sóból-* over time, Figure 10 is "easier" to read, but one should keep in mind that the earlier decades deal with very little data.

---

6. The present frequency findings do not correspond to the earlier ones for the overall distribution of the semantic range of *-sóból-*, but they do for the changes in overall occurrence frequencies through time. Observe, though, that only three snapshots were taken in the earlier study (for the 1900s, 1960 and 2010s), while each decade is sampled in the present more-detailed study.

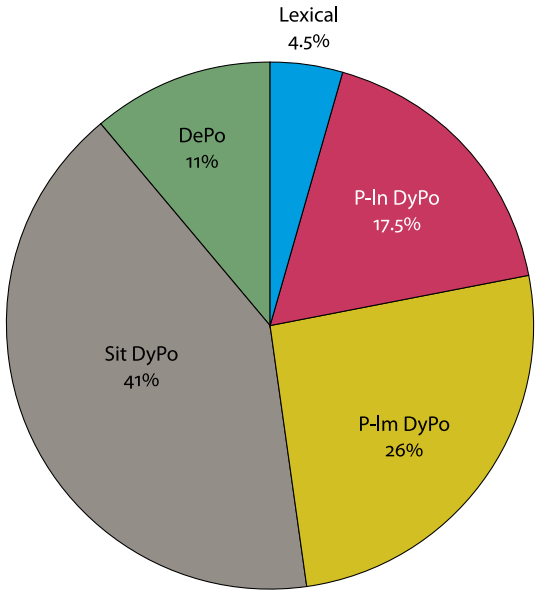

Figure 7. Overall distribution of the uses of *-sóból-*

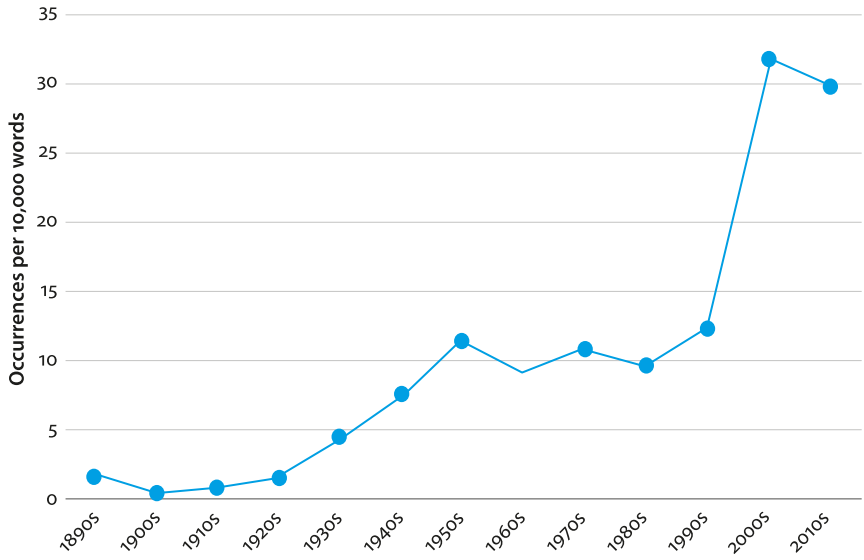

Figure 8. Overall occurrence frequency of *-sóból-* through time

The share of lexical meanings of *-sóból-* is seen to be significant in the earlier decades but reduces drastically over time. In the 2000s and 2010s, these meanings

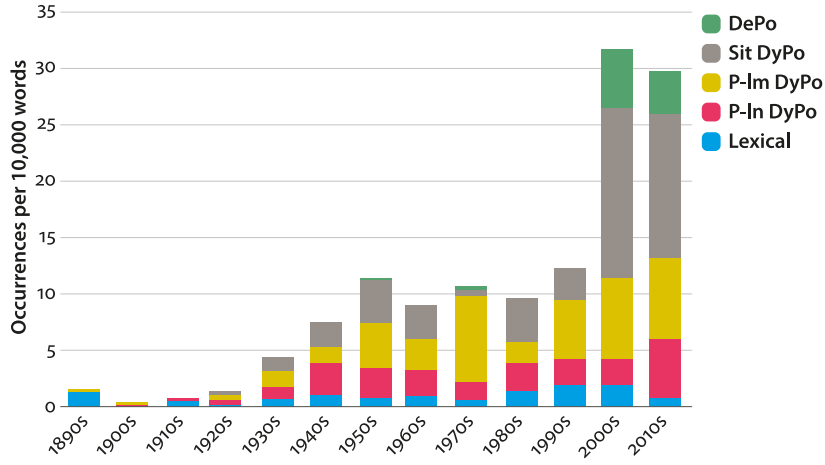

Figure 9. Distribution of the actual occurrences of the uses of *-sóból-* through time

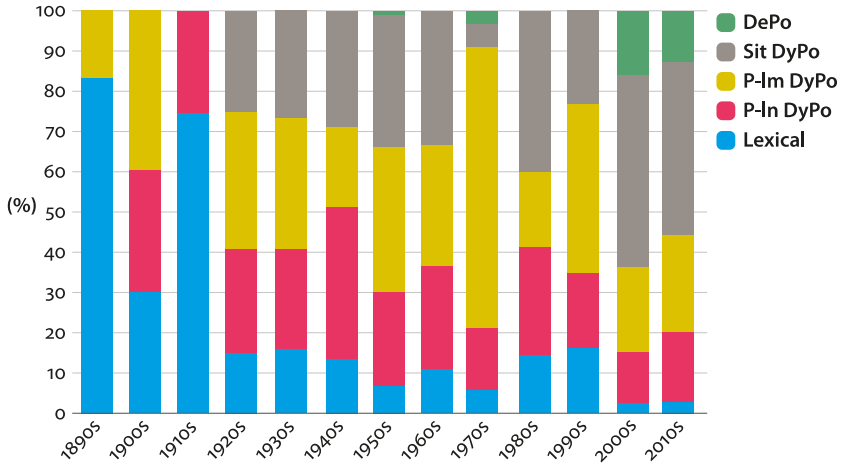

Figure 10. Percentage distribution of the uses of *-sóból-* through time

have become insignificant. Participant-inherent and participant-imposed dynamic possibility uses are significant throughout the decades and as a share of all the uses they remain relatively stable. Situational dynamic possibility uses enter the semantic range of *-sóból-* during the 1920s and these are seen to be increasing over time. Although traces of deontic possibility uses are visible in the 1950s and 1970s, strictly speaking, *-sóból-* is meaningfully associated with these uses only starting with the 2000s.

6. An account of *-andi-*

In their analysis of the modal verbal prefix *-andi-* on the basis of the same 4-million-word corpus, Kawalya et al. (2018a) note that the overall use of this affix has been constant through time and that it has also, just like *-sóból-*, become strongly associated with modality in recent times. In earlier periods, *-andi-* was mainly a conditional marker expressing counterfactuality and hypotheticality, while it was infrequently involved in the expression of modality, more specifically deontic necessity. However, from around the 1940s, the semantic range of *-andi-* extended from conditional meanings and deontic necessity to also include epistemic possibility. The distributional analysis of the verbal prefix *-andi-*, based on Kawalya et al. (2018a), is summarised below in order to complete the picture of the modal subdomain of possibility.

6.1 Distributional analysis of *-andi-*

The verbal prefix *-andi-* is found to express both conditional meanings (counterfactuality and hypotheticality) as well as modal meanings (deontic necessity and epistemic possibility). As shown in Figure 11, 49% of the uses express conditional meanings, with counterfactuality contributing 31% and hypotheticality 18%. Modal meanings contribute 46%, with deontic necessity taking 37% while epistemic possibility takes only 9%. The remaining 5% are cases of *-andi-* expressing both counterfactuality and deontic necessity.

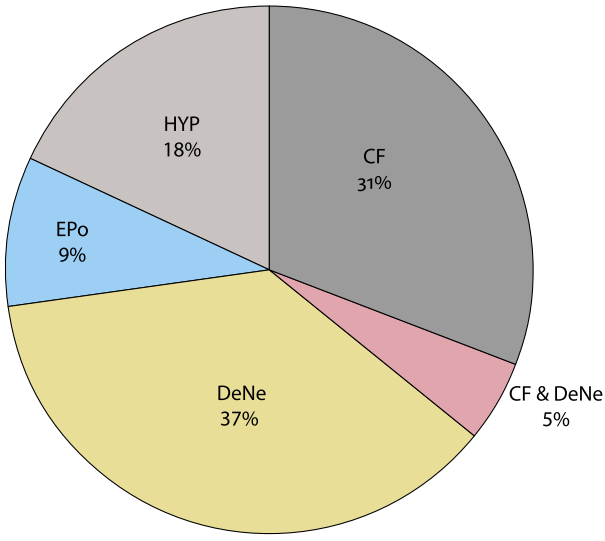

Figure 11. Overall distribution of the uses of *-andi-*

6.2 Diachronic distributional analysis of *-andi-*

Overall, *-andi-* occurs 3,998 times in the 4-million-word corpus. While also oscillating through time, the frequency averages 9.9 times for every 10,000 words, as may be seen from Figure 12. The distribution of the various uses of *-andi-* may be seen in Figures 13 and 14.

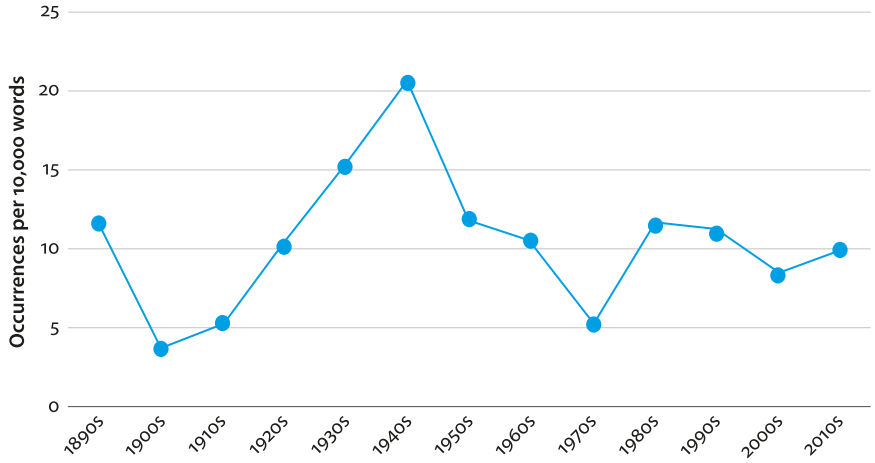

Figure 12. Overall occurrence frequency of *-andi-* through time

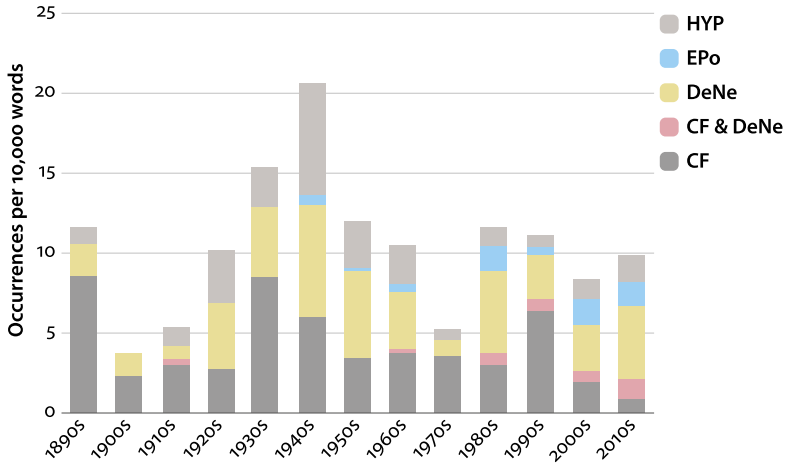

Figure 13. Distribution of the actual occurrences of the uses of *-andi-* through time

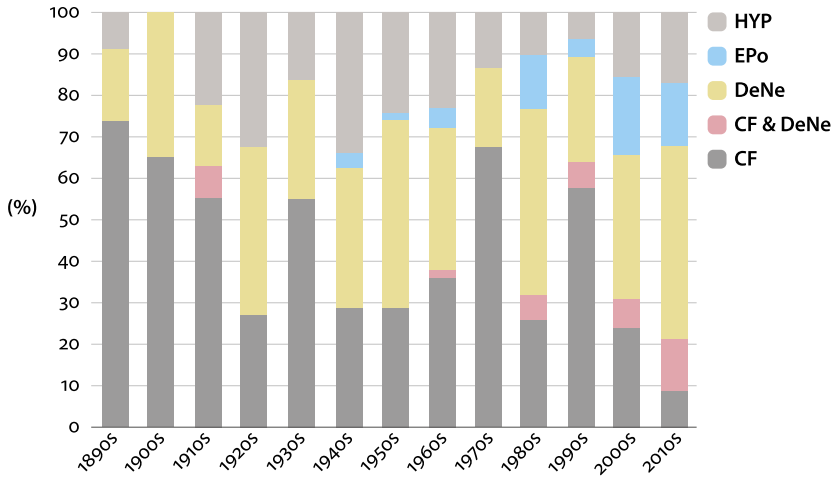

Figure 14. Percentage distribution of the uses of *-andi-* through time

From Figures 13 and 14, it may be seen that in the earlier decades, *-andi-* was mostly involved in the expression of conditional meanings, viz. counterfactuality and hypotheticality. Modal uses, in this case deontic necessity and epistemic possibility, have become more prominent over time, with epistemic possibility uses only first sighted in the 1940s and increasingly attested up to the 2010s, save for the 1970s where no occurrence is attested. In other words, in contrast to the two other markers discussed here, *-andi-* is only involved in the expression of the more subjective categories of deontic and epistemic modality without any association whatsoever with the more objective category of dynamic modality. Moreover, it is the only marker that straddles the dividing line between the modal subdomains of possibility and necessity, at least since the 1940s when we have the first empirical evidence for the appearance of epistemic possibility in the semantic range of *-andi-* along with deontic necessity already attested since the 1890s.

7. Discussion

From Figures 2, 7, and 11, which show percentages of the overall functional load of the respective possibility markers *-yînz-*, *-sóból-* and *-andi-*, we can conclude that over the last 130 years, *-yînz-* has had the highest functional load dedicated to possibility, i.e. expressing possibility (in all instances in which it occurs in double-verb constructions) in 97.5% of its total occurrences in the whole corpus, closely followed by *-sóból-* with 95.5%. Having only 9% of its attestations devoted to pos-

sibility, the prefix *-andi-* can be considered as Luganda's most marginal potential marker.

With 4.5% of its overall occurrences dedicated to lexical usages, *-sóból-* does so about 1.6 times more than *-yînz-* with only 2.5%. Moreover, the 2.5% for *-yînz-* include cases where the verb expresses a meaning that is ambiguous between its original lexical meaning and participant-inherent dynamic possibility. We can also deduce from Figures 3 and 7 that all three types of dynamic possibility take a much higher share of the functional load of *-sóból-* than they do for its near-synonym *-yînz-*. In fact, they do so twice more for *-sóból-* than for *-yînz-*. The share of deontic possibility, on the other hand, is two times larger for *-yînz-* than it is for *-sóból-*. Comparing Figures 3 and 7, we can also deduce that epistemic possibility is overwhelmingly expressed by *-yînz-*; in fact, its percent share is six times more than that of *-andi-*. Unlike all other types, epistemic possibility is not conveyed by *-sóból-*. Conversely, apart from epistemic possibility, no other possibility type is expressed by *-andi-*. This shows, therefore, that the more objective dynamic possibility meanings dominate the functional domain of *-sóból-*, while the more subjective attitudinal meanings (deontic and epistemic possibility) are mostly associated with *-yînz-*. The verbal prefix *-andi-* is only used with the most subjective possibility type, viz. epistemic possibility.

When *-sóból-* and *-yînz-* are compared to one another in respect of their dynamic possibility uses over time (i.e. the sum of participant-inherent, participant-imposed and situational uses), as shown in Figure 15, *-sóból-* is clearly seen to be taking over from *-yînz-* from the mid-20th century onwards. Its dynamic uses rise to 26 occurrences (for each 10,000 words) in the past two decades, compared to those of *-yînz-* which drop to only 3 occurrences in the same period. It remains unclear, however, whether the increase in the expression of possibility meanings for all the markers studied in this paper signifies an overall increase in the need to express possibility over the last 130 years. Our assumption is that either there has been, in recent years, an increase in registers that require possibility meanings more, such as health and inspiration, which make a big part of our corpus for the 2000s, or that the need to express possibility has not increased per se, but that possibility in the late 19th century is likely to have been expressed through other possibly less grammaticalised devices that were not the focus of this paper.

Turning to the lexical uses, Figure 16 shows that the lexical uses of *-yînz-* become ever less frequent until they completely disappear by the 1960s. Lexical uses of *-sóból-* are still attested in the language and have even slightly grown over time. For both verbs, however, these uses remain very limited.

From Figure 5 we see that the use of *-yînz-* to express deontic possibility has been rather stable through time, apart from an overrepresentation in the

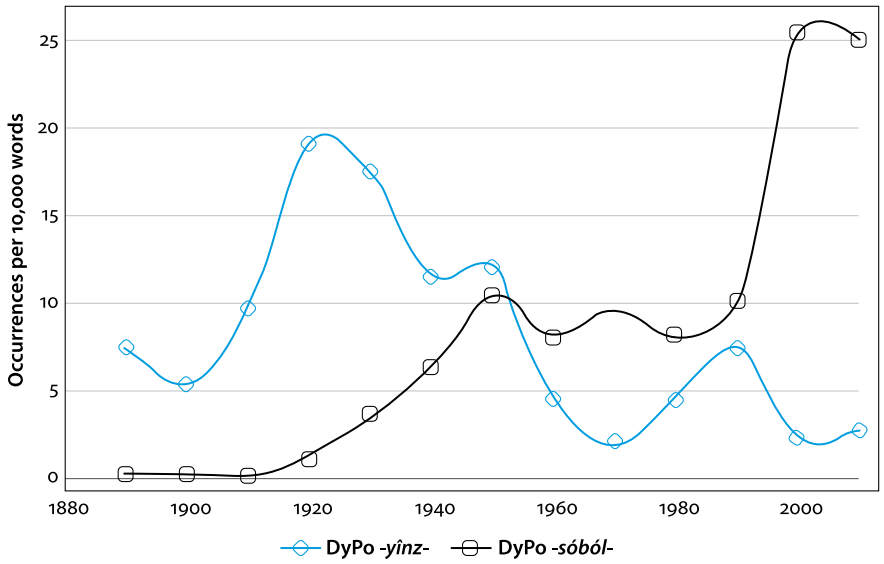

Figure 15. Occurrence frequencies of the dynamic possibility uses of *-yînz-* and *-sóból-* through time

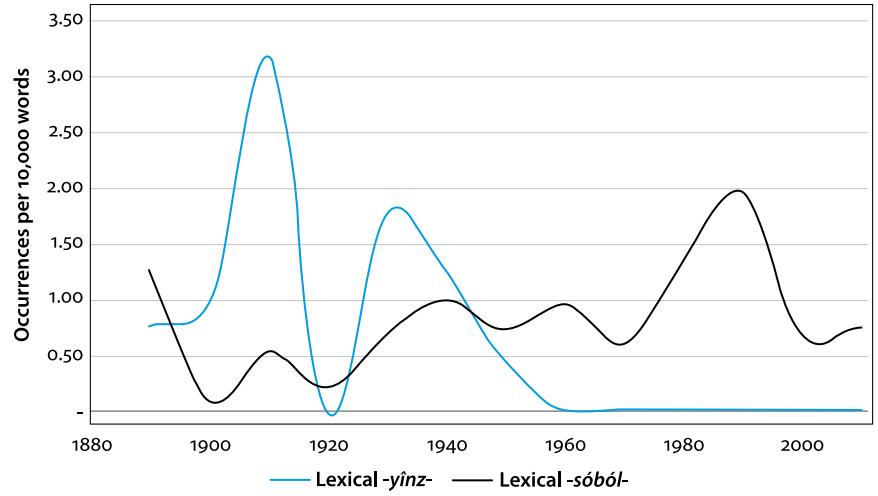

Figure 16. Occurrence frequencies of the lexical uses of *-yînz-* and *-sóból-* through time

1950s, while Figure 9 shows that the use of *-sóból-* to express this possibility type becomes significant only during the past two decades. Furthermore, Figure 5 shows that epistemic possibility also has been rather stably expressed by *-yînz-* through time, while Figure 13 shows that *-andi-* increasingly acquires this mean-

ing from around the 1940s onwards. In other words, neither the entrance of *-sóból-* into the subcategory of deontic possibility since the 1960s nor the appearance of *-andi-* in the subcategory of epistemic possibility since the 1940s had a similar impact on the modal use of *-yînz-* as the rise of *-sóból-* in the subcategory of dynamic possibility.

All of the above indicates that in addition to *-yînz-* losing its lexical meanings, its dynamic possibility uses are also almost disappearing. Simultaneously, *-sóból-* still maintains its lexical uses while at the same time increasing its modal uses, having acquired deontic possibility uses in the recent decades. Furthermore, this implies that in present-day Luganda, *-yînz-* can be regarded as the most developed modal marker as it ceased expressing lexical meanings while still covering the entire possibility spectrum (though only archaically as for dynamic possibility), unlike *-sóból-* which is still associated with lexical usage while not yet covering the entire possibility spectrum. The potential prefix *-andi-* only expresses epistemic possibility and deontic necessity.

8. Conclusions

In this corpus-driven analysis, we have studied the interaction of the three main Luganda possibility markers *-yînz-*, *-sóból-* and *-andi-*. Our analyses have shown that all the categories of possibility are associated with the modal verb *-yînz-*, while one of the categories, epistemic possibility, is not expressed by *-sóból-*. When *-andi-* is used to express modality, it is only found for epistemic possibility and deontic necessity uses. Furthermore, it was shown that the more objective dynamic possibility meanings are mostly expressed by *-sóból-*, while the more subjective ones, deontic and epistemic possibility, are mostly expressed by *-yînz-*. This shows that *-yînz-*, being more semantically diversified and mostly associated with the more subjective modality types (which are more central to modality than their objective counterparts) is more established as a modal marker than both *-sóból-* and *-andi-*.

Diachronically, although both *-yînz-* and *-sóból-* were involved in the expression of lexical meanings in the first half of the 20th century, *-yînz-* has not been associated with its original lexical usage since the 1960s. Furthermore, although all the more objective dynamic possibility meanings are expressed by both *-yînz-* and *-sóból-* throughout the past 13 decades, *-yînz-*'s role in the dynamic possibility subdomain appears to be more significant than that of *-sóból-* during the first half of the 20th century and greatly reduces in the second half. The role of *-sóból-* in the expression of dynamic possibility becomes important during the second half of the century. The more subjective possibility types are mainly expressed

by *-yînz-*, with *-sóból-* acquiring deontic possibility since the 2000s, while *-andi-* acquires epistemic possibility only around the 1940s.

This implies that *-yînz-* has been significantly associated with possibility meanings longer than *-sóból-*, whose association with possibility becomes significant only around the mid-20th century. Thus, one can tentatively conclude that *-yînz-* is historically more established as a marker of modality and is, therefore, likely to have been used to express modal meanings in earlier stages of the language when *-sóból-*, *-andi-* and probably other possibility markers, were still insignificant for these uses.

The increase of the involvement of *-yînz-* and *-sóból-* in the expression of the more subjective meanings over time implies that they are both undergoing a process of subjectification, from expressing the more objective meanings to expressing the more subjective ones. This also applies to *-andi-* which acquires, and becomes increasingly involved in the expression of, the more subjective epistemic possibility meanings over time. However, *-yînz-* appears to be more subjectified than *-sóból-*, as it already expresses all possibility types, while *-sóból-* not only became meaningfully associated with deontic possibility just recently but is also yet to express epistemic possibility.

Furthermore, it was shown that *-yînz-* expressed lexical meanings in those cases where it was used autonomously, without a second verb, and whenever it clearly expressed modal meanings, it was used in combination with another verb. Our diachronic analysis has shown that in present-day Luganda, *-yînz-* can no longer be used autonomously in a clause; it must, at all times, be followed by a second verb. The autonomous uses of *-sóból-* are still attested in present-day Luganda, although these are also infrequent. This implies that the two verbs are undergoing a process of grammaticalisation from a full verb to an auxiliary, which must always be used with a second verb. This grammaticalisation of *-yînz-* and *-sóból-* can be said to be at different stages, with the former being at a more advanced stage than the latter. Furthermore, like *-sóból-* (Kawalya et al., 2014: 81), this grammaticalisation of *-yînz-* in the structural domain seems to be correlated with subjectification in the semantic domain, since a reduction in their autonomous use goes hand in hand with an increase in their involvement in the expression of the more subjective modal uses.

With this global corpus-driven account of semantic change within the modal subdomain of possibility, we have shown the potential of corpus linguistic methodology for the study of Bantu diachronic semantics generally and modality more specifically without having to rely on comparative language data. At the same time, with Luganda's written tradition only starting in the 1880s, we need to admit the fact that the available text corpus is not old enough to trace back the beginnings of the interaction between the modal markers *-yînz-*, *-sóból-* and

-andi-. This absolute time barrier can only be overcome by complementing the present study with more traditional historical-comparative research on the closest Great Lakes Bantu relatives of Luganda.

Funding

The research presented in this article was supported by the Special Research Fund of Ghent University, through a doctoral scholarship awarded to the first author, with the other authors as supervisors.

References

Anderson, G. D. S. (2006). *Auxiliary Verb Constructions*. Oxford University Press. https://doi.org/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*. Longmans, Green and Co.

Bastin, Y. (1985). *Les relations sémantiques dans les langues bantoues* [Semantic Relations in the Bantu Languages]. Académie royale des sciences d'outre-mer.

Bastin, Y., Coupez, A., & de Halleux, B. (1983). Classification lexicostatistique des langues bantoues (214 relevés) [Lexicostatistical classification of the Bantu languages (214 records)]. *Bulletin des séances de l'Académie royale des sciences d'outre-mer*, *27*(2), 173–199.

Bostoen, K. (2008). Bantu spirantization: Morphologization, lexicalization and historical classification. *Diachronica*, *25*(3), 299–356. https://doi.org/10.1075/dia.25.3.02bos

Bostoen, K., & Bastin, Y. (2016). Bantu lexical reconstruction. In *Oxford Handbooks Online*. Oxford University Press.

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. https://doi.org/10.3406/aflin.2012.1005

Chesswas, J. D. (1963). *The Essentials of Luganda* (3rd ed.). Oxford University Press.

Cole, D. T. (1967). *Some Features of Ganda Linguistic Structure*. Witwatersrand University Press.

De Pauw, G., de Schryver, G.-M., & van de Loo, J. (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)* (pp. 85–92). European Language Resources Association.

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. https://doi.org/10.3406/aflin.2010.989

Devos, M. (2008). The expression of modality in Shangaci. *Africana Linguistica*, *14*, 3–35. https://doi.org/10.3406/aflin.2008.1048

Fleisch, A. (2008). The reconstruction of lexical semantics in Bantu. *Sprache und Geschichte in Afrika*, *19*, 67–106.

Kawalya, D., Bostoen, K., & de Schryver, G.-M. (2014). Diachronic semantics of the modal verb *-sóból-* in Luganda: A corpus-driven approach. *International Journal of Corpus Linguistics*, *19*(1), 60–93. https://doi.org/10.1075/ijcl.19.1.03kaw

Kawalya, D., de Schryver, G.-M., & Bostoen, K. (2018a). From conditionality to modality in Luganda (Bantu, JE15): A synchronic and diachronic corpus analysis of the verbal prefix *-andi-*. *Journal of Pragmatics*, *127*, 84–106. https://doi.org/10.1016/j.pragma.2018.01.011

Kawalya, D., de Schryver, G.-M., & Bostoen, K. (2018b). Reconstructing the origins of the Luganda (JE15) modal auxiliaries *-sóból-* and *-yînz-*: A historical-comparative study across the West Nyanza Bantu cluster. *South African Journal of African Languages*, *38*(1), 13–25.

Kawalya, D., de Schryver, G.-M., & Bostoen, K. (2019). A corpus-driven study of the expression of necessity in Luganda (Bantu, JE15). *Southern African Linguistics and Applied Language Studies*, *37*(4), 361–381. https://doi.org/10.2989/16073614.2019.1697624

Le Veux, H. (1917). *Premier essai de vocabulaire luganda–français d'après l'ordre étymologique* [First attempt of a Luganda – French vocabulary following an etymological order]. Imprimerie des Missionnaires d'Afrique (Pères Blancs).

Mberamihigo, F. (2014). *L'expression de la modalité en kirundi: Exploitation d'un corpus électronique* [The Expression of Modality in Kirundi: Exploitation of an Electronic Corpus]. Unpublished PhD, Université libre de Bruxelles / Ghent University.

Mberamihigo, F., de Schryver, G.-M., & Bostoen, K. (2016). Entre verbe et adverbe: Grammaticalisation et dégrammaticalisation du marqueur épistémique *umeengo/umeenga* en kirundi (bantou, JD62) [Between verb and adverb: Grammaticalisation and degrammaticalisation of the epistemic marker *umeengo/umeenga* in Kirundi (Bantu, JD62)]. *Journal of African Languages and Linguistics*, *37*(2), 247–286. https://doi.org/10.1515/jall-2016-0011

Mulira, E. M. K., & Ndawula, E. G. M. (1952). *A Luganda–English and English–Luganda Dictionary* (2nd revised edition of the dictionary compiled by Kitching & Blackledge in 1925). Society for Promoting Christian Knowledge.

Murphy, J. D. (1972). *Luganda–English Dictionary*. The Catholic University of America Press.

Nuyts, J. (2006). Modality: Overview and linguistic issues. In W. Frawley, E. Eschenroeder, S. Mills, & T. Nguyen (Eds.), *The Expression of Modality* (pp. 1–26). Walter de Gruyter.

Schadeberg, T. C. (2002). Progress in Bantu lexical reconstruction. *Journal of African Languages and Linguistics*, *23*(2), 183–195. https://doi.org/10.1515/jall.2002.011

Scott, M. (2019). *WordSmith Tools* (Version 7) [Computer software]. Lexical Analaysis Software. https://lexically.net/wordsmith/downloads/

Snoxall, R. A. (1967). *Luganda–English Dictionary*. Clarendon Press.

Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*. John Benjamins. https://doi.org/10.1075/scl.6

Tramutoli, R. (2015). 'Love' encoding in Swahili: A semantic description through a corpus-based analysis. *Swahili Forum*, *22*, 72–103.

Traugott, E. C. (1989). On the rise of epistemic meanings in English: An example of subjectification in semantic change. *Language*, *65*(1), 31–55. https://doi.org/10.2307/414841

van der Auwera, J., & Diewald, G. (2012). Methods for modalities. In A. Ender, A. Leemann, & B. Wälchli (Eds.), *Methods in Contemporary Linguistics* (pp. 121–142). De Gruyter. https://doi.org/10.1515/9783110275681.121

van der Auwera, J., & Plungian, V. A. (1998). Modality's semantic map. *Linguistic Typology*, *2*(1), 79–124. https://doi.org/10.1515/lity.1998.2.1.79

Appendix A. Abbreviations and symbols

APPL	applicative
AUGx	augment of class x
CAUS	causative
CF	counterfactuality
CONN	connective
DEMa	proximal demonstrative
DEMb	medial demonstrative
DeNe	deontic necessity
DePo	deontic possibility
EPo	epistemic possibility
FV	final vowel
HYP	hypotheticality
IPFV	imperfective
IRR	irrealis
Lexical	lexical meaning
Lexical^{+}	meaning which is ambiguous between lexical and participant-inherent dynamic possibility
LOCx	locative of class x
N	homorganic nasal
NEAR_FUT	near future
NEG	negative
NPx	nominal prefix of class x
Ø	null morpheme
OPx	object prefix of class x
P-In DyPo	participant-inherent dynamic possibility
P-Im DyPo	participant-imposed dynamic possibility
PASS	passive
PERS	persistive
PFV	perfective
PL	plural
POSSx	possessive of class x
POT	the potential verb *-yînz-*
PPx	pronominal prefix of class x
PRS	present
QUOT	quotative
REL	relative
REM_PST	remote past
SBJV	subjunctive
SG	singular
Sit DyPo	situational dynamic possibility
SPx	subject prefix of class x

Appendix B. Online supplementary material

Supplementary material, including raw frequencies, is available.

Address for correspondence

Gilles-Maurice de Schryver
BantUGent – UGent Centre for Bantu Studies
Ghent University
Blandijnberg 2
Ghent 9000
Belgium
gillesmaurice.deschryver@UGent.be
https://orcid.org/0000-0001-7272-9878

Co-author information

Deo Kawalya
School of Languages, Literature and Communication
Makerere University
kawalyad@gmail.com

Koen Bostoen
BantUGent – UGent Centre for Bantu Studies
Ghent University
koen.bostoen@UGent.be
https://orcid.org/0000-0003-2284-6165

A diachronic corpus-driven study of the expression of possibility in Luganda (Bantu, JE15)

Deo Kawalya, Koen Bostoen, and Gilles-Maurice de Schryver

ONLINE SUPPLEMENTARY MATERIAL

Supplementary Material 1: Brief overview of definitions and typologies for modality as found in the literature

In agreement with Lyons (1977: 452), Palmer (1986: 14-15) attempts to define modality in terms of the expression of "the speaker's attitude or opinion" or as referring to "all the non-propositional elements of a sentence", or still, as being "concerned with the status of the proposition that describes the event" (Palmer 2001: 1). As pointed out by Nuyts (2006: 6), however, this definition is very broad and it even includes categories like tense and aspect. For that reason, Nuyts (2006: 1) proposes a narrower definition of modality, where the notion refers to a semantic subfield within the wider TAM domain. Bybee *et al.* (1994: 176) apply a diachronic aspect to Palmer's definition: following Bybee (1985), they suggest that a real understanding of modality would emerge from a study of a set of "diachronically related functions". Hence they define modality as the "grammaticization of speakers' (subjective) attitudes and opinions" (Bybee *et al.* 1994: 176). On the other hand, Traugott & Dasher (2001: 50) define modality in terms of the "relativization of the validity of sentence meanings to a set of possible worlds" or the expression of "a perspective that considers the possibility of things being otherwise than they are" (see also Kiefer 1994: 2515). We stick to Nuyts (2006) because it is narrower.

Supplementary Material 2: Raw frequency facts that underlie Figures 5, 6, 9, 10, 13 and 14

Raw frequency facts that underlie Figure 5 (expressed as occurrences per 10,000 words)

Meaning	1890s	1900s	1910s	1920s	1930s	1940s	1950s	1960s	1970s	1980s	1990s	2000s	2010s
Lexical	0	0	0	0	0	0.4	0.4	0	0	0	0	0	0
Lexical+	0.8	1	3.2	0	1.8	0.8	0	0	0	0	0	0	0
P-In DyPo	4.8	3.1	5.3	4.4	9.3	4.2	1.7	2.6	0.4	1.5	2.6	0.2	0.2
P-Im DyPo	2	1.4	2.8	5.8	3.1	1.7	3.5	0.4	0.1	0.3	0.2	0.2	0.2
Sit DyPo	0.8	1	1.8	9.2	5.3	5.8	7	1.6	1.7	2.8	4.7	2	2.4
DePo	0.5	3.1	1.8	5.8	2.7	3.7	18.7	2.9	2.2	1.6	1.9	3.1	3.9
EPo	2.3	1	1.1	7.3	7.1	7.1	9.1	4.4	2	5.4	7.3	7.8	9.2
SUM	**11.2**	**10.6**	**16**	**32.5**	**29.3**	**23.7**	**40.4**	**11.9**	**6.4**	**11.6**	**16.7**	**13.3**	**15.9**

Raw frequency facts that underlie Figure 6 (expressed as relative frequencies)

Meaning	1890s	1900s	1910s	1920s	1930s	1940s	1950s	1960s	1970s	1980s	1990s	2000s	2010s
Lexical	0	0	0	0	0	1.7	1	0	0	0	0	0.0	0
Lexical+	7.1	9.4	20.0	0	6.1	3.4	0	0	0	0	0	0.0	0
P-In DyPo	42.9	29.2	33.1	13.5	31.7	17.7	4.2	21.8	6.3	12.9	15.6	1.5	1.3
P-Im DyPo	17.9	13.2	17.5	17.8	10.6	7.2	8.7	3.4	1.6	2.6	1.2	1.5	1.3
Sit DyPo	7.1	9.4	11.3	28.3	18.1	24.5	17.3	13.4	26.6	24.1	28.1	15.0	15.1
DePo	4.5	29.2	11.3	17.8	9.2	15.6	46.3	24.4	34.4	13.8	11.4	23.3	24.5
EPo	20.5	9.4	6.9	22.5	24.2	30	22.5	37	31.3	46.6	43.7	58.6	57.9
SUM	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**

Raw frequency facts that underlie Figure 9 (expressed as occurrences per 10,000 words)

Meaning	1890s	1900s	1910s	1920s	1930s	1940s	1950s	1960s	1970s	1980s	1990s	2000s	2010s
Lexical	1.26	0.1	0.53	0.21	0.68	1	0.73	0.96	0.6	1.35	1.96	0.69	0.76
P-In DyPo	0	0.1	0.18	0.35	1.09	2.82	2.66	2.32	1.67	2.58	2.31	4.14	5.29
P-Im DyPo	0.25	0.13	0	0.48	1.43	1.5	4.11	2.73	7.53	1.84	5.16	6.56	7.18
Sit DyPo	0	0	0	0.35	1.16	2.16	3.75	3.01	0.6	3.81	2.84	15.19	12.84
DePo	0	0	0	0	0	0	0.12	0	0.36	0	0	5.18	3.78
SUM	**1.51**	**0.33**	**0.71**	**1.39**	**4.36**	**7.48**	**11.37**	**9.02**	**10.76**	**9.58**	**12.27**	**31.76**	**29.85**

Raw frequency facts that underlie Figure 10 (expressed as relative frequencies)

Meaning	1890s	1900s	1910s	1920s	1930s	1940s	1950s	1960s	1970s	1980s	1990s	2000s	2010s
Lexical	83.4	30.3	74.6	15.1	15.6	13.4	6.4	10.6	5.6	14.1	16.0	2.2	2.5
P-In DyPo	0.0	30.3	25.4	25.2	25.0	37.7	23.4	25.7	15.5	26.9	18.8	13.0	17.7
P-Im DyPo	16.6	39.4	0.0	34.5	32.8	20.1	36.1	30.3	70.0	19.2	42.1	20.7	24.1
Sit DyPo	0.0	0.0	0.0	25.2	26.6	28.9	33.0	33.4	5.6	39.8	23.1	47.8	43.0
DePo	0.0	0.0	0.0	0.0	0.0	0.0	1.1	0.0	3.3	0.0	0.0	16.3	12.7
SUM	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**

Raw frequency facts that underlie Figure 13 (expressed as occurrences per 10,000 words)

Meaning	1890s	1900s	1910s	1920s	1930s	1940s	1950s	1960s	1970s	1980s	1990s	2000s	2010s
CF	8.6	2.4	3	2.8	8.5	6.0	3.5	3.8	3.6	3	6.4	2	0.9
CF&DeNe	0	0	0.4	0	0.0	0.0	0	0.2	0	0.7	0.7	0.6	1.2
DeNe	2	1.3	0.8	4.1	4.4	7.0	5.4	3.6	1	5.2	2.8	2.9	4.6
EPo	0	0	0	0	0.0	0.7	0.2	0.5	0	1.5	0.5	1.6	1.5
HYP	1	0	1.2	3.3	2.5	7	2.9	2.4	0.7	1.2	0.7	1.3	1.7
SUM	**11.6**	**3.7**	**5.4**	**10.2**	**15.4**	**20.7**	**12**	**10.5**	**5.3**	**11.6**	**11.1**	**8.4**	**9.9**

Raw frequency facts that underlie Figure 14 (expressed as relative frequencies)

Meaning	**1890s**	**1900s**	**1910s**	**1920s**	**1930s**	**1940s**	**1950s**	**1960s**	**1970s**	**1980s**	**1990s**	**2000s**	**2010s**
CF	74.1	64.9	55.6	27.5	55.2	29.0	29.2	36.2	67.9	25.9	57.7	23.8	9.1
CF&DeNe	0.0	0.0	7.4	0.0	0.0	0.0	0.0	1.9	0.0	6.0	6.3	7.1	12.1
DeNe	17.2	35.1	14.8	40.2	28.6	33.8	45.0	34.3	18.9	44.8	25.2	34.5	46.5
EPo	0.0	0.0	0.0	0.0	0.0	3.4	1.7	4.8	0.0	12.9	4.5	19.0	15.2
HYP	8.6	0.0	22.2	32.4	16.2	33.8	24.2	22.9	13.2	10.3	6.3	15.5	17.2
SUM	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**

References

Bybee, J. L. (1985). *Morphology: A Study of the Relation between Meaning and* Form. John Benjamins.

Bybee, J. L., Perkins, R., & Pagliuca, W. (1994). *The Evolution of Grammar: Tense, aspect, and modality in the languages of the world*. The University of Chicago Press.

Kiefer, F. (1994). Modality. In R. E. Asher & J. M. Y. Simpson (Eds.), *The Encyclopedia of Language and Linguistics* (pp. 2515–2520). Pergamon Press.

Lyons, J. (1977). *Semantics*. Cambridge University Press.

Nuyts, J. (2006). Modality: Overview and linguistic issues. In W. Frawley, E. Eschenroeder, S. Mills & T. Nguyen (Eds.), *The Expression of Modality* (pp. 1–26). Walter de Gruyter.

Palmer, F. R. (1986). *Mood and Modality*. Cambridge University Press.

Palmer, F. R. (2001). *Mood and Modality. Second edition*. Cambridge University Press.

Traugott, E. C., & Dasher, R. B. (2001). *Regularity in Semantic Change*. Cambridge University Press.