Long Default and Short Indefinite Noun Forms in Kupsapiiny: Synchronic Usage and Diachronic Development

Kazuhiro Kawachi

Keio University

Abstract: In Kupsapiiny, a Southern Nilotic language of Uganda, nearly all common nouns in the singular or the plural have two forms depending on the presence vs. absence of a suffix. This study examines how long and short common noun forms (with and without the suffix, respectively) occur in discourse by means of Dryer's (2014) Reference Hierarchy. It shows that short forms are only infrequently used for indefinites, with their mandatory use restricted to two types of indefinites, semantically nonspecific indefinites involving negation and true-predicate nominals, whereas long forms serve as frequently used, default forms, which occur not only in definite situations, but also – in fact, more commonly than short forms – in most indefinite situations. The present study hypothesizes that long forms have semantically generalized from forms marking definiteness, i.e., the speaker's assumption about the identifiability or accessibility of the referent to the hearer (Givón 1995, 2001), to those indicating the speaker's assumption about the accessibility of the referent (type) to the hearer (or sometimes the speaker themself) with the result that their context extended from definite situations to include most indefinite situations, as reported about the diachronic development of definite articles in other languages (e.g., Greenberg 1978). It speculates that the context extension for long forms was brought about by associations between referents in discourse based on the Gricean maxim of relation as well as subjectification. Other types of factors than the speaker's assumption about the referent (type) accessibility and indefiniteness in the choice between long and short forms (lexical, constructional, contextual, and generational differences as well as special uses of short forms for definites) are also described. This study also points out that the Kupsapiiny noun system as a whole contradicts the grammatical form–frequency correspondence hypothesis (Haspelmath 2021).*

Key words: (in)definiteness, Dryer's Reference Hierarchy, form–frequency non-correspondence, grammaticalization, Kupsapiiny (Kalenjin, Southern Nilotic, Uganda)

1. Introduction

Almost all common nouns in Kupsapiiny (ISO 639-3: kpz), a Southern Nilotic language of Uganda, each have long and short forms. The present study analyzes data on the ranges of the use of these noun forms in this language. It shows that while long forms occur exclusively in definite situations, they also commonly occur in most indefinite situations, whereas short forms are strictly limited in usage range to indefinites. In fact, long forms are used far more frequently than short forms even in indefinite situations. Accordingly, long forms are default forms, which occur virtually irrespective of definiteness, and short forms are indefinite forms, which are only infrequently used in indefinite situations. This study investigates the factors in the use of the two types of noun forms in Kupsapiiny, and hypothesizes about the diachronic development of long forms.

The rest of Section 1 provides a brief overview of grammatical properties of Kupsapiiny relevant to the present study. Section 2 gives a short outline of Dryer's (2014) Reference Hierarchy, and draws up a synopsis of the treatment of the distinction between long and short noun forms in Kupsapiiny and other languages in the Kalenjin branch of the Southern Nilotic group in the literature. Section 3 sketches out the set of forms that Kupsapiiny common nouns take and their morphological types (Section 3.1). It then describes the use of Kupsapiiny long and short noun forms for situations of different degrees of definiteness on the Reference Hierarchy, proposes the major factors (Section 3.2), and depicts other factors in their use (Section 3.3). It also presents data on the frequency of long and short noun forms in conversations (Section 3.4). Section 4 formulates a possible hypothesis about the diachronic development of long forms from definite forms to default forms – specifically, it surmises how the sort of context where long forms were used has extended from definite situations to include indefinite situations as well, and what instigated such a change (Section 4.1). It then speculates on the relations between the different factors in the use of long and short forms described in Sections 3.2 and 3.3 (Section 4.2). Section 4 also points out that the greater frequency of long forms than short ones can be analyzed as exhibiting a form–frequency non-correspondence (Section 4.3). Section 5 concludes the paper.

Kupsapiiny (with the second "p" pronounced as [b]; also spelled Kupsapiny or Kupsabiny) belongs to the Kalenjin branch of the Southern Nilotic group of the Nilotic language family, and is spoken by the Sebei people in the Sebei region (bordering on Kenya) on the northern and western slopes of Mt. Elgon in Eastern

Uganda. According to the 2014 national census, the population of the Sebei is 289,456, and almost all of them speak Kupsapiiny.

Kupsapiiny morphology is agglutinative, though it often exhibits a certain degree of fusion. Nouns mainly use suffixes, and verbs can carry both prefixes and suffixes. This language shows mostly head-marking properties. Kupsapiiny distinguishes two cases – the nominative case and the absolute (object) case – with tone. The absolute case is functionally unmarked in that forms in this grammatical case occur in a wider range of environments than those in the nominative case. In this respect, the case-marking system of this language can be regarded as marked nominative (Dixon 1994, König 2006, 2008). Kupsapiiny has a relatively rigid VSO word order, though VOS occurs especially when the subject is a full noun and the object is a pronoun, when the subject is a third person and the object is first or second person, and when the subject is second person and the object is first person. Kupsapiiny is primarily a spoken language, but for the transcription of Kupsapiiny forms, the present paper makes use of the writing system developed by my main consultant, Chebet Francis, and his colleagues for Bible translation (Kupsapiiny Language Board & Kupsapiiny Language Development and Preservation Foundation 2017).[1]

2. Background
2.1. Dryer's Reference Hierarchy

Dryer (2014) explains how he conducted research on definite and indefinite articles for his chapters (Dryer 2005a, b, 2013a, b) of the *World Atlas of Language Structures* (WALS). He treats definite and indefinite articles in a broad sense and includes any morphosyntactic form expressing definiteness (for example, a noun affix) that occurs in noun phrases, though he excludes adnominal demonstratives.

Based on data on the use of definite and indefinite articles in various languages, Dryer proposes a typology of definite and indefinite articles, using the Reference Hierarchy (henceforth, RH) in (1), where five types of noun phrases are arranged in the order of the degree of definiteness, i.e., the extent of the speaker's assumption that the referent is identifiable to or accessible to (namely, mentally

---

[1] The consonant letters used in the writing system are: p, t, c, k, f, s, sy (ʃ in IPA), m, n, ny (ɲ in IPA), ŋ, r, l, w, and y (j in IPA). The vowel letters are: i, e, ē, a, ā, o, ō, and u, which are each doubled to represent their long counterparts. The three vowels with a macron, whose pronunciation involves the advancement of the tongue root resulting in the expansion of the pharyngeal cavity, appear to be the [+ATR] counterparts of those without it. The author provided morpheme and clitic boundaries and added information on zero morphemes. The writing system capitalizes sentence-initial letters, but the present paper does not. Also unlike the writing system, the present paper does not use a slash to indicate a verb form with the impersonal subject prefix, which is identical in form to the first-person plural prefix. Tone is distinctive in this language, though the writing system only employs a colon right before the beginning of a common noun (not of a proper noun or a pronoun) as used as the subject of a verb other than *miitē* 'exist'. For the purpose of the present study, only segmental transcriptions are sufficient, with the result that tones are not represented.

representable in or retrievable from the mind of) the hearer (Givón 1995: 350, 2001: 459), which hinges on discourse anaphoricity and pragmatic and semantic specificity.

(1) (i) *Anaphoric definites*: noun phrases that refer back in the discourse
   (ii) *Nonanaphoric definites*: noun phrases that do not refer back in the discourse, but whose referents the speaker and the hearer have shared knowledge about (e.g., *the sun*)
   (iii) *(Semantically and) pragmatically specific indefinites*: noun phrases whose referents are newly introduced into and are mentioned again later in the discourse (e.g., After *I went to this movie last night*, the movie is talked about.)
   (iv) *Pragmatically nonspecific but semantically specific indefinites*: noun phrases whose referents are newly introduced into and are not mentioned again later in the discourse (e.g., After *I went to a movie last night*, the movie is not mentioned later.)
   (v) *Semantically (and pragmatically) nonspecific indefinites*: noun phrases whose referent existence is not entailed (e.g., *John is looking for a unicorn.*)

Dryer (2014: e241) treats articles more strongly associated with noun phrases higher and those more strongly associated with noun phrases lower on the RH as definite and as indefinite, respectively; hence, if a language has both a definite and an indefinite article, the definite one has a usage range higher than the indefinite one. According to him, if a language has an article used for more than one type of noun phrase on the RH, the types of noun phrases for which the article is used are contiguous on the RH.

There are two types of nominals that Dryer (2014: e235–236) states he does not include in his RH, namely, true predicate nominals (e.g., *He is a teacher*) and generics. According to him, "[t]rue predicate nominals are most commonly coded the same way as semantically nonspecific indefinites, while generics are coded like either semantically nonspecific indefinites or nonanaphoric definites" (e236).

2.2. Descriptions of the two types of noun forms in Kalenjin languages in the literature[2]

Short and long noun forms in Kalenjin languages (other than at least Akie: König et al. 2015, 2020; discussed later in Section 4.2) have been described with different names in the literature. Some researchers (e.g., Tucker & Bryan 1962, 1964, 1966: 12, 139–142, 208–220, 456–457, 461, Creider & Creider 1989: 32, 168–169) use "primary" and "secondary" forms to describe short and long forms. Toweett (1979: 421) calls short and long forms in Kipsigis (Kipsikiis) "inclusives", which include "every member of the group", and "exclusives", which "exclude members of other" groups. Baroja (1989: 18–20) terms short and long forms in Pokot (Pökoot, Päkot;

---

[2] Long and short forms are presented in the reversed order when their morphological marking is discussed or when their morphological forms are given in contrast to each other.

also called Suk) "indefinite article forms" and "definite article forms", which he states make a distinction similar to the English definiteness distinction, though Beech (1911: 51) describes this language as lacking this distinction. Hollis (1909: 160–164) regards the suffix on long forms in Nandi "the article" without mentioning its meaning. The short and long forms in Kupsapiiny are called as such, as well as "thematic" and "paradigmatic" forms, by Montgomery (1966: 125–154), who labels the suffix on long forms "the paradigmatic case suffix", and are called "simple and determined forms" by O'Brien & Cuyers (1975: 54–58), respectively. For Sabaot (more specifically, the Koony dialect), the other variant of Kupsapiiny spoken on the Kenyan side of Mt. Elgon, Larsen (1991: 145) prefers "to follow Rottland (1981) in describing the secondary form as the "unmarked" form, since it is the form commonly used", also stating, "The primary form is best described as 'indefinite'."

In the same vein of Larsen's comments, reports in the literature (e.g., Tucker & Bryan 1962: 139, Creider & Creider 1989: 168–169, Zwarts 2003: 110–110) run that long forms are more commonly used than short forms in at least some Kalenjin languages.[3] The common use of long forms for a variety of apparently indefinite situations must have made most of the above-mentioned researchers who described these languages reluctant to use the term "definiteness" to refer to the distinction between the two types of noun forms, even though Rottland (1981: 9) asserts: "There is evidence that in PSN [Proto-Southern Nilotic] the secondary suffixes were in some way linked with the notion of definiteness." The questions are what principles underlie the distinction, and what triggers the frequent use of long forms as well as the restricted use of short forms for certain indefinite situations. The present study deals with these issues in Kupsapiiny.

3. Data

This study examines data that the author collected during his sixteen 3–5 week field trips in Kapchorwa, Sebei, in Uganda between 2009 and 2020 and has been collecting online since April 2020. The data comes from different sources including conversations, folk tales, true stories, and memory-based descriptions of Chafe's (1980) Pear Film. Although I checked with my Kupsapiiny native speaker consultants whether and how preferably the type of form used in a specific situation can be replaced by the other type of form, the present study mainly deals with

---

[3] According to Tucker & Bryan (1962: 139), who call short and long forms in the Pokot language and the Nandi-Kipsigis Kalenjin subgroup, in which they include Kupsapiiny, "primary (indefinite)" forms and "secondary (definite)" forms, "(i)n Päkot the Secondary suffix may be said to correspond in function to the definite article 'the' in English; in Nandi-Kipsigis, on the other hand, the Secondary form of the noun is the norm (corresponding to both 'the' and 'a(n)' of English), while the Primary form denotes that the noun is being used in a very general sense, or else adverbially." Rottland (1981: 9) also states that in Kalenjin, "the functional distinction between primary noun forms and secondary noun forms is either "unmarked : definite" (e.g., Päkot) or "indefinite : unmarked" (e.g., Nandi).

the usage data, keeping their comments to a minimum.[4]

3.1. Kupsapiiny common noun forms and types

Nearly all common nouns in Kupsapiiny (and a limited number of proper nouns: e.g., *Sapiin-ø-caak* (Sapiiny.person-PL-x) in (11)) have four forms, depending on the combination of number (singular vs. plural) and the presence vs. absence of a suffix ("suffix x", henceforth): short singular forms, long singular forms, short plural forms, and long plural forms (see Section 3.3.1 for exceptions). Most such nouns take the form of "stem–number(–x)", though they differ in how clearly they can be divided into separate morphemes. The allomorphs of the singular suffix are *-ø*, *-(C)V*, and *-in*, whereas the allomorphs of the plural suffix are *-ø*, *-o(o)*, *-((C) V(V))n*, *-V(V)s*, and *-Vy*. (Note that *-ø* is an allomorph of either the singular or plural suffix, as discussed shortly.) Suffix x takes different forms depending on number: *-(n)teet/-(n)tēēt*, *-(V(V))t*, *-(t(y))o/-(t(y))ō*, or *-ta* for the singular, and *-(C)V(V)k* or *-ka* for the plural.

Nouns with the four types of forms are of four types, depending on the morphological (un)markedness of their number forms, as shown in Table 1. In the first type [a] (the pluralization type), singular forms are morphologically unmarked, with the singular suffix being zero. In the second type [b] (the singularization type), plural forms are morphologically unmarked, with the plural suffix being zero. In the third type [c] (the equipollent type), both singular and plural forms are morphologically either marked or unmarked, and it is difficult to decide whether it is their singular or plural forms that are more (un)marked. In the fourth type [d] (the suppletive type), it is difficult to analyze the forms into morphemes. Nouns of each type do not ostensibly appear to form any semantics-based category (but see Section 4.1).

3.2. Use of Kupsapiiny long and short noun forms for situations of different degrees of definiteness

Table 2 gives a brief summary of the use of Kupsapiiny long and short noun forms on the RH ("#" means "not used"). The usage range of short forms satisfies Dryer's (2014: e241) characterization of that of forms with an indefinite article: 'if an article is restricted to the last three types on the RH, then it counts as an indefinite article'. However, long forms are used not only for definite situations, but also for most apparently indefinite situations; in fact, generally, long forms are more common than short forms even in most indefinite situations, as pointed out about

---

Table 1 Four types of nouns in Kupsapiiny

(Un)markedness	Gloss	short singular SG	long singular SG-X	short plural PL	long plural PL-X
[a] SG unmarked	'large basket'	kuluunku-ø	kuluunkuu-ø-t	kuluunkuu-n	kuluunkuu-n-ak
	'rain'	roop-ø	roop-ø-ta	roop-kwōōn	roop-kwōōn-ik
	'hole'	mwēēn-ø	mwēēn-ø-ēt	mwēēn-ōōs	mwēēn-ōōsy-ēk/ mwēēn-oon-ik
	'song, music'	tuum-ø	tuum-ø-to	tuum-wōōn	tuum-wōōn-ik
[b] PL unmarked	'rat'	mur-yaa	mur-yaa-nteet	mur-ø	mur-ø-eek
	'egg'	mākēy-ō	mākēy-ōō-ntēēt	mākēy-ø	mākēy-ø-iik
	'young woman'	kōōkōōn-nyōō	kōōkōōn-nyōō-ntēēt	kōōkōōn-ø	kōōkōōn-ø-ik
	'thief'	cōōr-iin	cōōr-iin-tēēt	cōōr-ø	cōōr-ø-iik
	'tree'	kēēt-ø	kēēt-ø-it	keet-ø/kēēt-iin	keet-ø-iik
[c] Both marked or unmarked	'mango'	muyēēm-ncō	muyēēm-ncōō-ntēēt	muyēēm-pēēn	muyēēm-pēēn-ik
	'Irish potato'	puryaasy-aa	puryaasy-aa-nteet	puryaas-iin	puryaas-iin-ak
	'fish'	purpur-yōō	purpur-yōō-ntēēt	purpur-uun	purpur-uun-ak
[d] Suppletive	'cow, cattle'	taany	teeta	toc/tuc	tooka/tuuka
	'girl, daughter'	tyēē	cēēpto	tiipin	tiipiik

Table 2 Use of long and short noun forms in Kupsapiiny on Dryer's Reference Hierarchy

Noun phrase type		Use of long and short forms	
		long forms	short forms
(i) Anaphoric definites		OK	#
(ii) Nonanaphoric definites			
(iii) Pragmatically specific indefinites		OK	OK
(iv) Pragmatically nonspecific but semantically specific indefinites			
(v) Semantically nonspecific indefinites	NEG	#	OK
	Other than NEG	OK	OK

other Kalenjin languages (see Section 2.2).[5]

The following discussion describes the use of Kupsapiiny long and short noun forms for the five types of noun phrases on the RH (see Sections 3.3.1–3.3.3 for various exceptions).

First, the use of long forms for the two types of definites on the RH is straightforward. The definites require the use of long noun forms.

(i) Anaphoric definites: For referents that have appeared in the previous discourse, only long noun forms are used. For example, the speaker of (2) had said immediately before this that her mother gave birth to two boys, and referred to the boys with *wēēr-ø-iik* (boy-PL-X), not *wēēr-ø* (boy-PL).

(2)   ...   yooto   kā-cēē-yiy   wēēr-ø-iik   mpo
            that.time   T.PST.3-IMPRS-produce   boy-PL-X   EMPH
      kaa=nyaaceek.
      home.SG.X=1PL.POSS
      '...   at that time, when the boys were born in our home.'
                                                      (Conversation 2016.8–6: E.C. & C.F.)

(ii) Nonanaphoric definites: Also for referents that have not appeared in the previous discourse, but that are identified by the speaker and the hearer based solely on their shared knowledge, long noun forms are used. For example, in (3), a portion of an utterance where the speaker talks about how heavily drunk people behave, the clothes appear for the first time in the discourse, but because both the speaker and the hearer know that 'the clothes' refers to those that the drunkards (the subject of (3)) wear in this context, the long form *sir-ōō-k* (central and eastern dialects)/ *sur-ōō-k* (western dialects) (clothes-PL-X), rather than the short one *sir-ōōy/sur-ōōy* (clothes-PL), is used.

(3)   ø-puur-ē   ŋuny,   ku-pa   sur-ōō-k   liin.
      PRS.3-fall-IPFV   ground   3-go.PL   clothes-PL-X   that.way

---

[5] Nevertheless, my middle-aged consultants seem to find the use of long forms for some of such indefinite situations less appropriate or less correct than that of short forms.

'They (drunkards) fall down, and the clothes (that they wear) become scattered (*lit.* go that way).'   (Conversation 2018.1–6: C.F. & S.S.)

For the three types of indefinites, short forms may be used, but long forms, in fact, are more commonly used, except that in the case of semantically nonspecific indefinites involving negation (and an additional type of indefinite discussed shortly), only short forms are used. In this language, pragmatic specificity, i.e., whether or not referents are mentioned again later in the discourse, does not affect the use of long and short noun forms.

(iii) Pragmatically specific indefinites: In (4), where the speaker mentions this friend for the first time in the conversation and continues to talk about her as the main topic, she uses the long form of the noun for 'friend'. (See also (11a).)

(4) a-nyoor-u   yaat   paatee coorw-ee-t,   ...
     1SG-get-1   DM   only   friend-SG-x
     'I just made a friend (*lit.* got the friend), ...'
                                      (Conversation 2016.8–6: C.F. & E.C.)

On the other hand, in (5) and (6), where the cows and the bicycle each appear for the first time in these stories and are mentioned later, the short forms of the nouns occur.

(5) ara   ŋoo   nyiin   nyēē   ø-miitē   ku-cōō-n-ii
     I.wonder   who   that.one REL.SG PRS.3-exist 3-come.SG-H-IPFV
     kaa=nyuu   āk   toc   cēē   ø-mii
     home.SG=1SG.POSS COM   cow.PL   REL.PL   PRS.3-exist
     ku-kwōr-uu?
     3-take/bring-IPFV.H
     'I wonder who that one is who is coming to my home with cows that they
     are bringing.' (*lit.* 'I wonder, who is that one who is coming to my home with
     cows that they are bringing?')   (Story 2015.7/8–6: A.K.)
(6) ... ku-cōō-n-ti-kēēy   leek-wee-t ake   nyēē
         3-come.SG-H-IPFV-REFL   child-SG-x certain.SG REL.SG
     ku-pōōnt-ōō-n-uu   paskiili-ø   ...
     3-have-AL-H-IPFV.H   bicycle(ENG)-SG
     '... a certain child who came along by bicycle (*lit.* had a bicycle as they came
     along) was coming ...'   (Pear Story #12: S.B.)

(iv) Pragmatically nonspecific but semantically specific indefinites: The speaker of (7), who introduces a stick for the first time into the story, and does not mention it after this anymore, uses the long form of the noun for 'stick'.

(7) ku-nam kōrōk-ø-tō   kule, ā-miir-ō,   ...
     3-grab stick-SG-x   QUOT   1SG-hit-o:1SG
     'He (in this context) (*lit.* 3) grabbed a stick to hit me (*lit.* saying I hit me), ...'
                                                       (Story 2018.1–1: C.P.)

On the other hand, in (8) and (9), where the people (the specific people who were killed by the president) and the tractors (the specific tractors that referents of the subject used) appear for the first time in the conversation and are not mentioned later, the short forms of the nouns are used.

(8) ø-pākāc-ē   piic   Amiin-i   …
     PRS.3-kill-IPFV   person.PL   Amin-EMPH
     'President Amin was killing people …' (Conversation 2016.8–8: C.F. & E.C.)

(9) … cēē-kwōō-n-ōō-t-ēē   kurakut-aan.
        1PL-come.PL-EP-AL-TH-VAFW   tractor-PL
     '… we repeatedly came using tractors.' (Conversation 2016.8–8: C.F. & E.C.)

(v) Semantically nonspecific indefinites: The use of long and short noun forms somewhat differs in the two types of cases of semantically nonspecific indefinites, one of which involves negation and the other does not.

When a noun denoting a semantically nonspecific indefinite is in the scope of negation, its short form usually occurs and its long form hardly ever does.[6] For example, in (10) and (11b), where the non-existence of a bridge and that of a gun are implicated, the short form of the noun for 'bridge' and that for 'guns' are used, respectively.

(10) … mā-wō=cēē-pun-t-ēē   taraac-a   soo.
         NEG-go=1PL-pass-TH-VAFW   bridge-SG   DM
     '… we did not go and pass a bridge that way.'
                                    (Conversation 2016.8–8: C.F. & E.C.)

(11) … ki-poont-o   pii=cu   (a) mōōt-ii-k   …   nteenee   aceek
         D.PST.3-have-3   person.PL=these gun-PL-X   but   1PL
         Sapiin-ø-caak   ku=cēē   ki-mā-cēē-pōōnt-ē
         Sapiiny.person-PL-X   COP=REL.PL   D.PST-NEG-1PL-have-1
         (b) mōōt-iin.
             gun-PL
     '(It turned out that) these people (were very tough and) had (a)guns, (and they had all the things that were enough to defeat us,) but we, the Sebei, were the ones who did not have (b)guns.'   (Story 2015.7/8–1: B.K.)

As the subject of the verb *puryo* 'not exist' or immediately after the quantifier *puryo* 'no', a short form occurs to indicate the entailed non-existence of any instance of

---

[6] In three out of 12 instances in my conversation data (Section 3.4) where the existence of newly mentioned referents is negated (6 with *puryo* 'not exist; no'), long forms were used: (a) *ām-ø-iik* (food-PL-X), (b) *sur-ee-t* (cloth-SG-X), and (c) *roop-ø-ta* (rain-SG-X) in the contexts of (a) 'there was no food, either, in those days', (b) 'if you don't have even clothes' (Conversation 2016.8–6: E.C. & C.F.), and (c) 'when there is no rain' (Conversation 2016.8–7: C.F. & C.F.), respectively. (a) occurred with *puryo*, whereas (b) and (c) transpired with a negated verb. Notice that *ām-ø-iik* is usually used in its long form, and (b) and (c) occurred in subordinate clauses, where the non-existence of the referent is not asserted, but is presented as a condition or a future situation.

the specified referent, as in (12).

(12) ... kii-puryo   mpo   <u>suputaal-ø</u>, puryo <u>kiy</u>   ne.
     D.PST.3-not.exist   even   hospital-SG   no   thing.SG EMPH
     'There was not even a hospital, there was nothing.'
                                            (Conversation 2016.8–5: C.F. & E.C.)

For semantically nonspecific indefinites that do not involve negation, either long forms (e.g., (13)) or short forms (e.g., (14), (15)) are used.

(13) ø-mii   mpo   <u>piiko</u>   Kaapcēēsōōmpē coo   miyoot-ec,
     PRS.3-exist also   person.PL.X   Kapchesombe   REL.PL   bad-PL
     ku-yuu   coo   ø-pōōkiitōōs-i,   <u>cōōr-ø-iik-i</u>,   mpo
     3-be.like   REL.PL PRS.3-get.drunk-EMPH thief-PL-X-EMPH   even
     <u>pōōn-ø-iik</u>,   ø-miitē   Kaapcēēsōōmpē=ē?
     wizard-PL-X   PRS.3-exist Kapchesombe=Q
     'Are there also bad people like drunkards (those who get drunk), thieves, or even wizards in Kapchesombe?'   (Conversation 2016.8–7: C.F. & C.F.)

(14) ... nto   ø-miitē   wulēē <u>cwiiyōōt-iin</u>?
     or   PRS.3-exist where difference-PL
     '... or are there differences (*lit.* ... or is there where there are differences)?'
                                            (Conversation 2016.8–3: C.F. & C.E.)

(15) ki-cii-peet-ee   kārēēnk-ōō-k ākoy Puukwo wuliin   kele
     D.PST-1PL-go.PL-VAFW foot-PL-X   up.to Bukwo   that.way QUOT
     ā-cēēŋ-ōō-t-ii   <u>kar-iin</u>   cē=po   sukuulu-ø.
     1SG-look.for-AL-TH-IPFV.TH   money-PL   REL.PL=POSS school-SG
     'We went on foot that way up to Bukwo looking for school fees (*lit.* saying I am looking for school fees while going along that way).'
                                            (Conversation 2016.8–8: C.F. & E.C.)

We now turn to the two types of nominals mentioned in Section 2.1 that are not included in the RH. True predicate nominals are almost always rendered in short forms (see (35) for an exception). This applies to any of the three copula constructions (expressing 'A is B') in this language shown in (16). The subject (A) can be in any person in (16a), has to be in the third person in (16b), and has to be in the first or second person in (16c).

(16) a. A *ku*=B: the construction whose predicate the copula proclitic *ku*= attaches to
     b. B A: the construction that juxtaposes a predicate and a subject noun phrase
     c. *yiiku* A B: the construction that uses the verb for 'become' *yiiku*

When a generic statement is made with any of the three constructions containing a nominal predicate, the predicate noun is in its short form, as in (17), (18), and (19). (The subject pronoun omitted from each example is in parentheses.)

(17) (neeto) ta=ku=<u>kariiman-iin</u>   paatee …
     3SG   still=COP=young.man-SG   only
     'He (in this context) (*lit*. 3SG) is still just a young man ...'
                                        (Conversation 2018.1–6: C.F. & S.S.)

(18) nteenee   ntēē   ta=<u>surukan-ø</u>   (neeto).
     but   DM   still=strong.person-SG   3SG
     'But he (in this context) (*lit*. 3SG) is still a strong person.'
                                        (Conversation 2018.1–7: S.S. & C.F.)

(19) ā-yiik-u   (akweek)   <u>pērpēr-ø</u> kōt.
     PRS.2PL-become-2 2PL   fool-PL   very
     'You (PL) are great fools.'   (Conversation 2018.1–6: C.F. & S.S.)

For the other type of nominal not included in the RH, namely generics, long forms (e.g., (20)) or short forms (e.g., (21)) occur.

(20) ku-mmwow   kāāpōōmc-ø-iit kule, "ā-yōōm-ē   anii <u>piiko</u>."
     3-say   monster-SG-X   QUOT PRS.1SG-eat-IPFV 1SG person.PL.X
     'The monster said, "I eat people".'   (Story 2015.7/8–3: C.S.)

(21) ... yoo   ki-tā-cēē-ŋōōsy-ēē   <u>rēēk-o</u>   ...
         where   D.PST.3-still-IMPRS-grind-VAFW   hand.mill-SG
     '… to where people were still grinding using a hand mill ...'
                                        (Conversation 2018.1–1: C.F. & C.S.)

As seen so far, short forms occur in any type of indefinite situation as indefinite forms, but semantically nonspecific indefinite nominals involving negation and true predicate nominals are exclusively restricted to short forms. Both types of contexts have one property in common – in either case, the noun phrase makes no explicit reference to any specific entity; it is strictly non-referential.[7] A semantically nonspecific indefinite involving negation negates the existence of an entity in question, and a true predicate nominal describes the property of the referent of a subject noun phrase. On the other hand, while definite situations exclusively require long forms, which also commonly appear in most indefinite situations, serving as default forms, which are practically neutral as to definiteness. Therefore, unlike the usage of short forms, which is limited to indefinite situations, the usage of long forms cannot be characterized in terms of definiteness – the major factor in their use as default forms in discourse seems to be the speaker's assumption about the accessibility of the referent or the type thereof of the noun phrase to the hearer in the context, as discussed later in Section 4.1, which addresses the question of how the two types of forms came to exhibit such a default vs. indefinite contrast in usage-range distribution, and develops a hypothesis that this contrast developed from the definiteness distinction.

---

[7] According to one of the reviewers, noun phrases in classifier languages can show the following property: "noun phrases with a classifier tend to have a definite interpretation, while the use of bare noun phrases is limited to generic, indefinite, or non-referential contexts".

3.3. Other factors motivating and hindering the use of Kupsapiiny long and short noun forms

As shown in Section 3.2, long common noun forms are used as default forms, and short ones occur only in indefinite situations. However, these criteria are not completely decisive. There are other factors that affect the choice between long and short forms; some of them may appear idiosyncratic and inexplicable at first glance. This section describes such factors including lexical differences (Section 3.3.1), constructional differences (Section 3.3.2), two special uses of short forms for definite situations (Section 3.3.3), and a generational difference (Section 3.3.4).

3.3.1. Lexical differences

Different types of noun lexemes can exhibit distinct characteristics with respect to the use of their long and short forms, not necessarily conforming to the default vs. indefinite contrast.

First, there are nouns whose long forms are used and short forms are not used even for indefinites in my conversation and story data, though their short forms are possible in indefinite contexts and in the constructions that always use short forms (Section 3.2.2). They include at least: *maa-ø-ta* as used in the sense of 'fire' (rather than 'gun'), *yoom-ø-eet* 'wind', *mees-ee-t* 'table' (SWH), *mucēēl-ēēn-ik* 'rice' (usually PL), *pēēko* 'water' (PL only), *cēēko* 'milk' (PL only).[8] Such nouns are inanimate, and denote natural forces or daily necessities, which are so commonly found as to be likely to convert indefinite situations into those close to nonanaphoric definites.

Second, locational nouns, most of which solely have singular forms, have only short forms (e.g., *taay* 'front', *keelya* 'direction', *tarat* 'top, aboveness') or both short and long forms (e.g., *ŋuny-ø* − *ŋuny-ø-iit* 'ground, belowness', *tapan-ø* − *tapan-ø-ta* 'side'), whose short forms exclusively occur in my conversation and story data. (Hereafter, examples of short and long forms in pairs are each separated with a dash.)

Third, there are nouns for locations whose short forms are always used even for definites and long forms occur only in the constructions where long forms occur (Section 3.3.2). They include at least (only singular forms are listed here): *lakam-ø* − *lakam-ee-t* 'hill', *kusaaw-a* − *kusaaw-ee-t* 'field', *āyn-o* − *āyn-ee-t* (central and eastern dialects)/*ānny-o* − *ānny-ee-t* (western dialects) 'river', *tawun-ø* − *tawun-ii-t* 'town (ENG)', *tekeey-ø* − *tekeey-ee-t* 'veranda', *kusap-a* − *kusap-ee-t* (LUG)/*kanis-a* − *kanis-ee-t* (SWH) 'church', *sukuulu-ø* − *sukuuluu-ø-t* 'school (ENG)', *kwirir-i* or *kwirr-i* − *kwirir-i-t* or *kwirr-ii-t* 'bed', and *māpuusi-ø* − *māpuusii-ø-t* 'prison'.

There are also nouns that have number-neutral, locational forms in addition to short and long forms, and when they denote a location, their locational forms

---

[8] The Kupsapiiny words in this subsection for which no morpheme-by-morpheme glosses are provided are presented in the form of "NOUN.STEM-SG(-X)" (or "NOUN.STEM-PL(-X)" when specified as PL). Those without any morpheme boundary belong to the suppletive type [d] in Table 1.

occur, and neither short nor long forms are used. Such nouns are at least (in addition to locational forms, singular short and long forms are listed here): *mpareen* 'garden' (*mpar-ø – mpar-ø-eet*), *rōōtiin* 'banana plantation' (*rōōt-ø – rōōt-ø-iit*), *kāpēn* 'cave' (*kāpēn-nyō – kāpēn-ēē-t*), *kurkat* 'doorway' (*kurk-a – kurk-ee-t*), *wōōk* 'forest' (*wōōk-ē – wōōk-ēē-t*), and *araan* 'road' (rather than in the sense of 'clan' or 'time') (*ar-ø – ar-ø-eet*).[9]

Finally, kinship terms, which lack morphological plural forms, each have only single forms that appear to be short forms (e.g., *yēēyō* 'mother', *paapa* 'father', *kōōkō* 'grandmother', *maama* 'uncle', *yeeya* 'sibling', *ceepiya* 'sister').

3.3.2. Constructional differences

Either long or short forms may be used, depending on the type of modifier that modifies the noun or the type of enclitic or suffix that attaches to it.

[1] Possessive constructions: There are two possessive constructions with full-noun possessors. In one (possessum-*aap* possessor), possessum nouns, which carry the possessive suffix -*aap*, are always in their long forms (e.g., *kwēy-ø-ōōk-aap wēēr-ii-t* (shoe-PL-X-POSS boy-SG-X) 'the boy's shoes'). In the other construction (possessum *nyē=po* (SG)/*cē=po* (PL) possessor), which uses *nyē=po* (SG)/*cē=po* (PL) consisting of the relative marker *nyēē* (REL.SG)/*cēē* (REL.PL) and the possessive marker *po* (*lit.*, 'belong'), possessum nouns can be either in their long or short forms (e.g., *paskiilii-ø-t/paskiili-ø nyē=po piiko alak* (bicycle(ENG)-SG-X/bicycle(ENG)-SG REL.SG=POSS person.PL.X certain.PL) 'certain people's bicycle'; also (15)), with the use of short forms confined to indefinite situations.

There are three possessive constructions for pronominal possessors. In one, the possessive pronoun (e.g., *nyeemwaanii* 'mine (SG)'), which is a combination of the relative marker *nyēē* (REL.SG)/*cēē* (REL.PL) and a possessive pronominal enclitic (e.g., =*mwaanii* (1SG.POSS)), modifies long noun forms (e.g., *waar-wee-t nyeem-waanii* (goat-SG-X 1SG.POSS) 'my goat', *kaayta nyaaceek* (home.SG.X 1PL. POSS) 'our home'). There is also a contraction of this construction (e.g., *waar-wee= mwaanii, kaa=nyaaceek* in (2)), where the pronominal enclitic attaches to a stem truncated at its end. There is a further construction with the pronominal suffix, which also attaches to the truncated stem of a long form (e.g., *waar-wee-nyuu* (goat-SG.X-1SG.POSS) 'my goat').

[2] Demonstratives: There are two types of demonstratives, adnominal demonstrative words and demonstrative suffixes, which each make a distinction between three degrees of distance of an entity from the speaker, "proximal", "medial", and "distal". The adnominal demonstrative word, which is more emphatic, modifies a long noun form (e.g., *waarw-ee-t nyi* (goat-SG-X this) 'this goat', *maa-ø-ta nyiin* (fire-SG-X that.over.there) 'that fire over there'), whereas the demonstrative suffix attaches to short noun forms (sometimes with some sound modification such as vowel lengthening or epenthesis) (e.g., *waarw-a-a-ni* (goat-SG-LV-this) 'this

---

[9] The last noun in this list also has the senses of 'clan' and 'time', and when it is used in either of these senses (e.g., (27)), its locational form cannot be used.

goat', *ma-ø-a-niin/ma-ø-a-yiin* (fire-SG-LV-that.over.there) 'that fire over there'),
though they could also be analyzed as contracted from long forms with sound
modification.

[3] Indefiniteness markers: With *(=)ake* (SG)/*(=)alak* (PL) '(a) certain, (an)other'
(e.g., (6), (25), (33), (34)), which often serves as an enclitic, long forms are used.
Nevertheless, this indefiniteness marker as used in the sense of '(an)other' can
attach to short forms in indefinite situations, as in (22), which is a negative sen-
tence, where the existence of the referent of the noun phrase is negated with a
combination of the negative prefix on the verb and *ake* conveying 'not any other'.

(22) … mā-ø-cāmē   ŋētyuny <u>tyoony</u>   ake   tōkōl, …
       NEG-PRS.3-like lion.SG animal.SG other.SG all
       '… (the) lion does not like any other animal, …'   (Story 2015.7/8–7: C.W.)

The enclitic *=o/=a* (*=to* after a stem ending in *(i)in*) 'a certain, some (in an
affirmative declarative sentence), any (in a negative declarative sentence or an
interrogative sentence)' attaches to short forms, as in (23).[10]

(23) ø-mii   <u>ŋal-ø=a</u>   cēē   pērē nyii ā-tēēp-ēēn=aanii?
       PRS.3-exist word-PL=any REL.PL   want 2SG 1SG-ask-o:1SG=1SG
       'Are there any other things (*lit.* any words) that you (SG) want to ask me (*lit.*
       that you (SG) want, I ask me)?'   (Conversation 2016.8–2: C.F. & C.M.)

[4] Numerals: When accompanied by a numeral, long noun forms are usually used,
as in (24).

(24) ... ka-poont-o   <u>kārēēp-ōōn-ik</u>   sōmōk.
        T.PST.3-have-3 basket-PL-X   three
        '... he (in this context) (*lit.* 3) had three baskets.'   (Pear Story #2: C.M.)

Short forms are also possible in indefinite situations, though (25) is the only one
instance of the modification of a short form with a numeral in the Pear Story data.

(25) ... kii-mii   piiko   alak   cēē   kii-til-e
        D.PST.3-exist person.PL.X certain.PL REL.PL   D.PST.3-pass-IPFV
        cēē   kii-puntooy   <u>wēēr-ø</u> sōmōk ...
        REL.PL   D.PST.3-reach   boy-PL   three
        '… there were certain people who were passing by, and who were about three
        boys (*lit.* who reached three boys) …'   (Pear Story #7: S.S.)

[5] Adnominal interrogatives: The adnominal interrogatives *nee* 'what' (e.g., (26))
and *ŋōō* 'who' modify short noun forms that precede them, and *ānkōōnu* (SG)/
ānkōōcu (PL) 'which', *ata* 'how many', and *tyaa* 'how much' each modify long forms
that precede or follow them (e.g., (27), (28)) and also modify short forms that pre-
cede them (e.g., (29)). Short forms, which can occur in the cleft construction with

---

[10] Two of the allomorphs of this enclitic, *=o* and *=to*, are identical in form to, but are differ-
ent from those of the suffix X for singular nouns in Section 3.1.

one of these three interrogatives, are restricted to indefinite situations.

(26) … <u>kii-pōōrw-ōōn</u>   nee   kōt   cēē
    D.PST.3-disease-PL   what   very.much   REL.PL
 ki-cii-nyōōy-ē?
 D.PST.3-IMPERS-treat-IPFV
 '… what kinds of diseases were often treated?' (*lit.* 'The things that IMPERS treated very much (are) what diseases?')

(Conversation 2016.8–5: C.F. & E.C.)

(27) kā-ŋōt   <u>ar-ø-eet</u>   ata?
 T.PST.3-remain   road-SG-X   how.many
 'How many times (*lit.* roads) have the things remained (like that)?' (The subject, presumably *ŋal-ø-eek* (word-PL-X) 'things, situation (*lit.* words)', is omitted from right after *kā-ŋōt*.)   (Conversation 2017.1–1: A.S., A., & F.)

(28) ata   <u>nkuruuwēē-n-ik</u>   cēē   ø-māc-ē   nyii?
 how.many   pig-PL-X   REL.PL   2SG.PRS-want-2   2SG
 'How many (of the) pigs do you (SG) want?' (*lit.*, 'Is it how many pigs that you (SG) want'?)   (elicited on March 4, 2022: C.F)

(29) <u>nkuruuwēē-n</u>   ata   cēē   ø-māc-ē   nyii?
 pig-PL   how.many   REL.PL   2SG.PRS-want-2   2SG
 'How many pigs do you (SG) want?' (*lit.*, 'Is it how many pigs that you (SG) want'?)   (elicited on March 4, 2022: C.F)

[6] Vocatives: For most nouns, their short forms are used as vocatives (e.g., *coorw-aa* (friend-SG) 'Friend!', *kāā-nēēt-iin* (NMLZ-teach-SG) 'Teacher!', *leek-waa* (child-SG) 'Child!', *sukōōr-iin* (soldier-SG) 'Soldier!', *wōōrkōōy-ō* (prophet-SG) 'Prophet!'). Nevertheless, with a small number of nouns, perhaps nouns that describe a person with a certain property, their long forms are used as vocatives (e.g., *pōō-n-tēēt* (old.man-SG-X) 'Old man!', *cēēsy-ēē-t/cēēpyōōsy-ēē-t* (middle.aged.woman-SG-X) 'Middle-aged woman!').

3.3.3. Two special uses of short forms

As shown in Section 3.2, short forms are used mostly for indefinites. However, there are two types of situations where short forms are used for definite referents.

First, in folk tales, short forms may be consistently used to indicate the speaker and the hearer's shared inaccessibility of the referent. For example, in the stories in my data where animals appear as protagonists, the short forms of the nouns are consistently used for the animals throughout the story (e.g., (22)). Short forms used this way seem to function like nicknames, and such use makes the storyteller and the hearer establish shared inaccessibility as part of their common ground.

Second, short forms may be used for clearly definite situations to convey the speaker's pejorative attitude toward the referent. For example, before (30), the porridge, which is the topic of the story, had previously been expressed in its long form many times, but the protagonist uses its short form here to communicate the

nuance of 'mere porridge'.

(30) ... ku-miitē   ku-teep-ee-keey kule,   "ām nee   sa-wuny
        3-exist   3-ask-IPFV-RFL QUOT   LOC what   PURP.1SG-hide
    <u>musar</u>   paatee kōt   neenyi   kaa-lay?"
    porridge.PL only   until   DM   T.PST.1SG-burn
    '... she was asking herself, "Why had I been hiding mere porridge until I got
    burned?"'   (Story 2015.7/8–5: C.F.)

The lexeme-, construction-, and context-specific factors in the choice between long and short forms described in Sections 3.3.1–3.3.3 are further scrutinized in Section 4.2.

3.3.4. Generational difference

There are young speakers of Kupsapiiny who almost always produce only the long forms of most common nouns in their daily lives, and seldom use short forms. There are even those who cannot provide short forms of some nouns, especially in the plural, when they are asked to. This might be a contributing factor in the high frequency of long forms.

The author investigated the frequency of long and short noun forms in his Pear Story data, where 21 participants had been instructed to describe the events in the Pear Film based on their memory, in a way that they would tell their stories to those who had never seen the film.[11] The author found 204 instances in total where participants used common nouns with the long vs. short form distinction to describe the entities in (31) when they first appear in the film.[12] The number

---

[11] As one of the reviewers commented, there might be a potential methodological problem with the use of Pear Story data in general for studies on phenomena involving referent accessibility. Because participants talk about the events that they have seen in the film, they are already familiar with them at the time of speaking. Moreover, in telling their stories, the participants may presume that the experimenters already saw the film, despite the instructions to describe the events in a way that they would tell their stories to those who have never seen the film.

Admittedly, the Pear Film is likely not to elicit descriptions of semantically nonspecific indefinites – unlike the natural conversation data, our Pear Story data contains few descriptions of semantically nonspecific indefinites, only one of which is (34). The low frequency of the use of short forms for semantically nonspecific indefinites in the Pear Story data (1 instance: 12.5% of the use of short forms) as compared to that in the conversation data (13 instances: 61.9% of the use of short forms) seems to have contributed to the extremely low frequency of short forms in our Pear Story data.

[12] The following were left out from analysis: English nouns that have no long- vs. short-form distinction (e.g., *furuutis* 'fruit': Pear Story #6: C.J., #18: C.N., #19: A.J.) and noun phrases without a noun (Dryer 2004) including (a) those made up only of a noun modifier (e.g., *ake* 'another (one)' for the boy: Pear Story #18: C.N.) and (b) so-called headless relative clauses (e.g., *nyēē kā-sut-ōō-t-i kaar-ii-t* (REL.SG T.PST.3-ride-AL-TH-IPFV bicycle(SWH)-SG-X) 'the one who was riding a bicycle along that way' for the boy: Pear Story #10: A.S.).

immediately after each item in (31) indicates the number of participants who used a noun with its long vs. short form distinction when the item initially appeared in the video. However, there were 81 instances that used long forms in constructions where long forms are always required (Section 3.3.2), including 62, where long forms were accompanied by *(=)ake* (sg)/*(=)alak* (pl) '(a) certain, (an)other', with which the newly introduced human characters were often mentioned. These were all excluded from analysis, and the remaining 123 cases were analyzed as those where these entities were mentioned at their first appearance with nouns involving the choice between long and short forms unaffected by the construction; they are itemized by the numbers in square brackets in (31).

(31) 'rooster'/'hen': 5 [5], 'fruit'/'passion fruit'/'avocadoes'/'oranges'/'mangoes' (for 'pears'): 18 [12], 'man': 20 [3], 'tree': 12 [9], 'basket(s)': 17 [15], 'ladder': 4 [4], 'cloth' (for 'apron'): 7 [7], 'pocket': 3 [3], 'rope': 3 [3], 'goat': 18 [12], 'boy'/'child'/'young man': 17 [3], '(the boy's) bicycle': 19 [17], 'girl'/'child': 8 [0], '(the girl's) bicycle': 12 [7], 'stone': 10 [10], 'boys'/'children': 18 [1], 'hat': 13 [12]

Out of the 123 instances, 115 (93.5%) used long forms, whereas as few as 8 (6.5%) used short ones. The entities for which short forms were used are all non-human: 'passion fruit'/'avocadoes'/'oranges' (for 'pears') (3 participants), 'goat' (1 participant), '(the boy's) bicycle' (2 participants), '(the girl's) bicycle' (1 participant), and 'stone' (1 participant). (There is only one instance in the data where a short form occurred as the subject of a sentence, which is generally unlikely to be indefinite (Givón 2018 [1979]): 'stone' in the context of 'A stone knocked (him)' (Pear Story #5: S.M.).)

The participants' ages range from 20 to 80 with the average of 37.0. Since the overall frequency of long forms is so high, it seems to be difficult to find any substantial age difference in the frequency of their use. Nevertheless, a difference is evident between the average age of the speakers who produced any short form (one or two short forms at most) and that of those who did not at all – out of the 21 participants, the average age of the six who produced them was 53.0, whereas that of the fifteen who did not produce any short forms was 27.4.

3.4. Frequency of Kupsapiiny long and short noun forms in conversations
The present subsection furnishes data on the frequency of Kupsapiiny long and short noun forms in conversations to demonstrate how frequently long forms are used in indefinite situations. By using data from six natural conversations that he obtained in 2016, the author investigated 191 cases involving the choice between the long and short forms of nouns in introducing new referents.[13] The

---

[13] Excluded cases involve (i) the use of a noun as a predicate that describes the subject noun phrase whose referent has already appeared in the discourse, (ii) the use of a noun with no long vs. short form distinction (e.g., kinship terms, locational nouns, English nouns), and (iii) the use of a noun phrase without a noun (Dryer 2004).

following types of descriptions were not taken into account for the analysis: (a) the apparently non-anaphoric definite use of long forms (with different degrees of palpability) (51 instances), for example, that for a body part of a previously mentioned human referent (8 instances), (b) the use of the short form of the noun for a location (Section 3.3.1) (3 instances), (c) the use of the locational form of a noun, rather than its long or short form (Section 3.3.1) (11 instances), and (d) the obligatory use of long or short forms due to one of the constructions that always use long forms (26 instances) or short forms (9 instances) (Section 3.3.2). Out of the remaining 91 instances where long or short nouns were used for the introduction of genuinely new referents without any lexical or constructional restriction, 70 instances (76.9%) used long forms, whereas 21 (23.1%) used short forms with 8 for semantically specific indefinites and 13 for semantically nonspecific indefinites (9 involving negation and 4 not involving it).

Therefore, long forms not only cover a wider range of noun phrase types on the RH, but also are more frequent in usage for indefinites in conversations than short forms.

4. Analysis and discussion

This section addresses issues involving Kupsapiiny long and short noun forms. Section 4.1 speculatively reflects upon how the context for the usage of long forms has extended from definite situations over the course of their development and hypothesizes what caused the context extension. Section 4.2 further speculates on how the lexeme-, construction-, and context-specific uses in Sections 3.3.1–3.3.3 are related to their basic uses described in Section 3.2 in view of their synchronic contrast and their hypothesized development. Section 4.3 points out that Kupsapiiny long and short noun forms systematically display a form–frequency non-correspondence.

4.1. Context extension for long forms in the course of their grammaticalization from definite to default forms

The usage extension of definite markers to indefinite situations in different languages has been reported in the literature (e.g., Greenberg 1978, 1981, 1991, Hawkins 2004: 82–93). Greenberg hypothesizes that definite articles (stage I) develop into non-generic, as opposed to generic, or specificity markers (stage II), which then end up being meaningless nominal markers or gender or noun-class markers (stage III) (see Schuh 1983 for Chadic examples).[14] At stage II, where all nouns still have forms with and those without the article, forms with it are used not only for definite but also for indefinite situations, and the article is highly grammaticalized to the extent that the choice between the two types of forms is

---

[14] Greenberg (1981, 1991: 302–305) describes the prefix *k-* in Eastern Nilotic languages as a stage III article. Its cognate in Kupsapiiny seems to occur only as part of male proper names (e.g., *kiprotic*) or the nominalizing suffix (e.g., *kāā-pāt–iin* (NMLZ-dig-SG) 'farmer'), and this prefix has nothing to do with the suffix -x in question.

determined by the construction – for example, articleless forms may be used only as the object of a negative generic sentence or as a nominal predicate. This seems closely akin to the situation observed in the use of Kupsapiiny long and short noun forms.[15] In fact, the Kupsapiiny noun system might be about to go beyond stage II to stage III – as Greenberg hypothesized about stage III, there seems to be at least one semantically motivated noun class that has been formed: nouns ending in *-(n)teet/-(n)tēēt* (SG) tend to mean an occupation, a person of a particular propensity, a creature, or a food item including a vegetable, a fruit, and a grain. A further indication of this developmental stage of the Kupsapiiny noun system is presented below in terms of the usage distribution of long and short forms.

As described in Section 3.2, short noun forms in Kupsapiiny always occur only for the two types of nominals, namely semantically nonspecific indefinite nominals involving negation and true predicate nominals, which are excluded from the use of long forms. These types of nominals possess one property in common – they are strictly non-referential. As Givón (2001: 442) states, "[n]ominals *may* be interpreted non-referentially only if they are under the scope of some non-fact modality" (i.e., some irrealis assertion or negation assertion modality). The two types of nominals are under the scope of such non-fact modalities causing a non-referring interpretation: semantically nonspecific indefinite nominals involving negation are under the scope of negation, and true predicate nominals, or attributive nominal predicates, "are inherently habitual and thus timeless", and "[t]he habitual tense/aspect is ... a sub-mode of irrealis" (Givón 2001: 449). Hence, non-referentiality, which requires the use of short forms and precludes that of long forms, is more restrictive than the non-entailment of the existence of the referent of a noun phrase as the defining property of semantically nonspecific indefinites on the RH, in which case the referent of the noun phrase may or may not exist.

On the other hand, Kupsapiiny nouns can take long forms unless the non-existence of the referent of the constituted noun phrase is entailed or they are non-referential predicates. Hence, in addition to definites and semantically specific indefinites, they cover some semantically nonspecific indefinites, having gone beyond Greenberg's stage II, which is defined in terms of semantic specificity.[16] Therefore, through the grammaticalization of long forms from definite forms, such "context extension" (Heine & Kuteva 2007, Narrog & Heine 2021) from definite situations can be hypothesized to have taken place in a way that indefinite situa-

---

[15] In fact, although he does not furnish concrete data, Greenberg (1978: 58), who touches on Southern Nilotic languages, states the following: "The NANDI-KUPSIGIS and PO-KOT (SUK) groups have genetically cognate suffixes for singular and plural, without gender classification. In POKOT these function like a definite article, while in NANDI-KUPSI-GIS" they are in the stage of non-generic article" – according to Tucker & Bryan (1962, 1964, 1966), Kupsapiiny is closer to Nandi and Kipsigis than to Pokot. See also footnote 3.
[16] Hawkins (2004) also states that at a late stage of the grammaticalization of a definite marker, noun phrases with it "simply assert existence" (2004: 86) before it becomes completely unrelated to definiteness.

tions also came to be included in their usage range, as in Greenberg's scenario.[17] As compared to context extension, a formal change is unlikely to occur until later, with the result that long forms are still longer than short forms, which are less frequent.

The major factor in the current use of long forms as default forms in discourse is not definiteness but is of a more relaxed quality – so relaxed as to be the speaker's assumption that the hearer (or even at least the speaker themself) can access information about what the referent of the noun phrase is like. With the use of long forms, in definite situations, the speaker can readily assume that like themself, the hearer can easily identify the referent, but even in most indefinite situations, the speaker may assume that the hearer also knows the referent or (if not the exact referent, at least) the type of the referent from the context or from their experience or world knowledge. For example, in (32), which occurs at the very beginning of a story, the storyteller uses the long form *kumny-at-eet* (honey-SG-X), because he knows, and believes that the hearer also knows, what type of honey the characters are looking for, regardless of the actual existence of such honey.

(32) kii-ŋeet   pēsy-ēē-t=ake,   Sumōōtwō ākoo   Ceemutaay
     D.PST.3-start   day-SG-X=certain   Sumootwo COM   Chemtai
     ku-pa   ku-cēēŋaat   <u>kumny-at-eet</u>   wuliin   po   wōōk.
     3-go.PL   3-look.for   honey-SG-X   there   belong/POSS   forest.LOC
     'A long time ago (*lit.* Starting on a certain day), Sumootwo and Chemtai
     went to look for honey there in the forest.'   (Story 2015.7/8–2: C.F.)

Accordingly, the notion of definiteness, i.e., the speaker's assumption about the identifiability or accessibility of the referent to the hearer (Givón 1995: 350, 2001: 459), seems to have been semantically generalized (Heine & Kuteva 2007: 36) for long forms through their grammaticalization. First, the strict notion of "the identifiability or accessibility of the referent to the hearer" became lenient to be replaced by a milder notion of "the *accessibility* of the referent or the referent *type*", making it superfluous for the hearer to be able to identify the exact individual referent. Second, the accessibility of information about the referent (type) is not necessarily based on the speaker's objective assessment of the hearer's knowledge state, as in the case of definiteness, but is estimated by the speaker on the analogy of or even to be equivalent to the speaker's own familiarity (Christophersen 1939, Hawkins 1978).

There seem to be two semantic-pragmatic factors that contributed to the semantic generalization or desemanticization of long forms from erstwhile definite forms to default forms in Kupsapiiny.

First, since conversation participants are expected to follow Grice's (1975, 1989) maxim of relation, speakers usually talk about entities related to each other

---

[17] One of the reviewers points out that the "development from definite markers to general nominal markers (and even to non-referential markers) is well documented for Asian languages, as discussed in several chapters in Yap (2011)".

within a discourse, and the hearer also expects them to – the speaker knows that this helps the hearer associate referents with one another, and increases their accessibility to them, even if it does not make them precisely identifiable to them. Even an entity that is mentioned for the first time in the discourse can often be easily associated with another that has already appeared in the discourse and be accessible to the hearer. For such an entity, a long form can be used. For example, in (33), the referent of the noun phrase in question *laat-yee-t* (neighbor-SG-X) can easily be associated with that of another noun phrase *neeto* (3SG), even though the neighbor appeared in the story out of the blue. (In this example, the long form is preferred to the short form. In fact, my main consultant rejects the use of the short form *laat-yaa* (neighbor-SG) in this context, because the storyteller can assume that anyone including the woman in the story has a neighbor.) As such, the association makes the situation appear more like a nonanaphoric definite on the RH.

(33) ... ku-lim   neeto   <u>laat-yee-t</u>   ku-ŋōōlōōl-ōō-n-uu   āk
      3-hear   3SG   neighbor-SG-X   3-talk-AL-H-IPFV.H   COM
    ciit=aake.
    person.SG.X=certain.SG
    '... she (in this context) (*lit.* 3SG) heard a neighbor talking with someone as she came.'   (Story 2015.7/8–5: C.F.)

The second factor is subjectification, whereby "meanings become increasingly based in the speaker's subjective belief state/attitude toward the proposition" (Traugott 1989: 35) (also, Lyons 1982). Speakers very commonly talk about (types of) entities that they know well and may project their familiarity with them onto the hearer, assuming that they are also familiar with them as a member of the same speech community.[18] In the course of the development of long noun forms in Kupsapiiny, the notion of the accessibility of the referent (type) seems to have transformed from a hearer-basis to a speaker-centered one. Consequently, long forms may be used when the referent is familiar to the speaker, and the referent (type) is presumed to be also accessible to the hearer, if not identifiable to them, as in (32) and (33).

Nevertheless, there are differences among speakers as to their familiarity with specific entities. For example, in (11a), the storyteller used the long form of the noun for 'guns' *mōōt-iik* (gun-PL.X). He had spent most of his life in Bukwo, the eastern-most district of the Sebei region bordered by Kenya, where the people including him are familiar with guns, which the government had provided to them in the past. According to my consultants, speakers in other districts, including themselves, who are unfamiliar with guns, would rather use the short form *mōōt-*

---

[18] It seems to be more polite or nicer to the hearer if the speaker assumes that they (the hearer) share knowledge with themself (the speaker) or have the ability to identify the referent than they do not. Discourse may generally move in this direction because of the cooperative nature of communication or perhaps occasionally due to speech errors or speakers' misunderstanding.

iin (gun-PL).

When the speaker is not sure about what the described entity is, they can use the short form of a noun that might denote it. Thus, a fundamental baseline for the use of a long form is the speaker's knowledge. The speaker of (34), who does not know of what type of fruit the referents are, uses the short forms *mucuunkw-āān* (orange-PL) and *matuunt-aan* (passion.fruit-PL).

(34) ... a-kas   ciit=aake   nyēē   ka-laany   kēēt-i-t
     1SG-see   person.SG.x=certain.SG   REL.SG   T.PST.3-climb tree-SG-x
     tarat ø-mii   ku-put-ē   ...   mā-nkēt   le,   ø-mii
     top   PRS.3-exist 3-pick-IPFV   ...   NEG.1SG-know   CMPL   PRS.3-exist
     ku-put-ē   mucuunkw-āān   nto   matuunt-aan.
     3-pick-IPFV   orange-PL   or   passion.fruit-PL
     '... I saw a certain person who climbed to the top of the tree, picking ... I do not know if he (in this context) (*lit.* 3) was picking oranges or passion fruit.'
     (Pear Story #10: A.S.)

On the other hand, there are two instances of the use of a long noun form even as a true predicate nominal. One of them is (35), where the speaker, who knows the person in question (her boyfriend) well, uses the long form of the noun for 'thief, cheat (SWH)' to describe him, even though the hearer does not know him.

(35) muyōōyiin-ø-tēēt   nyiin   tōkōl, ...
     thief(SWH)-SG-x   that.one   totally
     'That one is a complete cheat (*lit.* thief).'
     (Conversation 2015.12–1: F.A. & C.F.)

This suggests that the use of long forms as default forms might have been further extended to the degree that they are on the verge of including true-predicate nominals.

4.2. Speculations about the relations between the basic uses of Kupsapiiny long and short noun forms and their lexeme-, construction-, and context-specific uses

The following discussion speculates on how the lexeme-, construction-, and context-specific uses of long and short forms in Sections 3.3.1–3.3.3, which are approximately summarized in Table 3, are related to their basic uses described in Section 3.2 in view of their synchronic contrast and their hypothesized development.

Table 3 Lexeme-, construction-, and context-specific uses of long and short forms

Long forms	Short forms
· nouns for natural forces or daily necessities · possessum nouns in possessive constructions (other than *nyē=po* (SG)/*cē=po* (PL) construction) · nouns with adnominal demonstrative · nouns with *(=)ake* (SG)/*(=)alak* (PL) '(a) certain'	· locational nouns · nouns for locations · kinship terms · vocative nouns · nouns in nickname use in stories · pejoratively used nouns · nouns with demonstrative suffix · nouns with enclitic *=o/=a* 'a certain, some, any' · nouns with adnominal interrogative 'what'/'who'
· nouns with *(=)ake* (SG)/*(=)alak* (PL) '(an)other' · nouns with adnominal interrogative 'which'/'how many'/'how much' · possessum nouns in possessive construction with *nyē=po* (SG)/*cē=po* (PL) · nouns with numeral	

The left and right columns of Table 3 list the types of noun lexemes or nouns in specific constructions or contexts whose long forms always or normally occur and those that invariably or customarily take their short forms, respectively. The bottom rows of Table 3 occupying both columns show nouns that can take their long or short forms according to the meaning or the context.

The default vs. indefinite distinction or its hypothetically original definite vs. indefinite distinction seems to be relevant to the distribution of noun types to some extent. Nouns denoting natural forces or daily necessities might be treated as non-anaphoric definites, which can easily be linked to any event participant in the conversational or story setting. Possessum nouns take their long forms in the possessive constructions (other than the one that uses *nyē=po* (SG)/*cē=po* (PL)) – possessum nouns generally tend to be definite (Haspelmath 1999), but unlike languages that do not bother to mark possessum nouns for definiteness due to economic motivation, Kupsapiiny seems to have been employing an explicit and uneconomical way in these possessive constructions. Nouns modified by an adnominal demonstrative are definite by definition, but this language also seems to have developed redundant definiteness marking for such nouns. (On the other hand, the demonstrative suffix in the right column economically attaches to apparently short forms.) The two kinds of adnominal interrogatives, which differ in the hearer's knowledge about the denoted noun, give rise to the different use of two types of forms – when 'what' or 'who' is used, the speaker has no knowledge of and asks questions about the identity of a hyponym or specific member of the denoted category denoted by a modified noun, which takes its short form, whereas with 'which', 'how many', or 'how much', both the speaker and the hearer may or may not have knowledge about the identity of the entity denoted by a noun that the adnominal interrogative modifies, resulting in the use of its long or short form. The use of short forms with the indefinite enclitic *=o/=a* 'a certain, some, any' shown

on the right column of Table 3 simply reflects the indefiniteness of the cliticized forms. However, plausibly as default forms, long forms are almost always used with the other indefiniteness marker *(=)ake* (SG)/*(=)alak* (PL) '(a) certain, (an)other'; nevertheless, short forms are used when the noun phrase is non-referential in the context of 'not any other'. Possessum nouns in the *nyē=po* (SG)/*cē=po* (PL) possessive construction as well as nouns with numerals can take their short forms only in indefinite situations. The nickname use and the pejorative use in the right column seem to come from one aspect of indefiniteness, which is a property of the referent or referent type unknown to the speaker (though the nickname use could also be attributed to the use of short forms like proper nouns).

On the other hand, indefiniteness seems irrelevant to the four items at the top of the right column (locational nouns, nouns for locations, kinship terms, and most vocative nouns), which either only have short forms or take their short forms even when they are undoubtedly definite or referential. They occur so frequently that they apparently deviated in their behavior from other common nouns, and function like grammatical morphemes or proper nouns. (In fact, locational nouns can function syntactically like prepositions.)

These types of nouns are used in kinds of situations where the referent is perceived as easily identifiable even without definiteness marking on the noun. Therefore, the present paper hypothesizes that these seemingly short forms had been, in fact, formerly long forms, which lost their final consonant in process of time with the result that they look like short forms. This is born out by an analogous phenomenon in another Kalenjin language without the long vs. short form distinction, Akie. According to König et al. (2015, 2020) and König (p.c. in May 2019 and January 2022), nouns in Akie have only forms that appear to correspond mostly to short forms in other Kalenjin languages (e.g., *chii* 'person', *láákwɛɛ* 'child'). Nevertheless, they can be hypothesized as long forms whose final consonant *t* or *k* was deleted in the past (e.g., *muŕantɛɛ* 'rat', *keeti* 'tree', *chááruɛ* 'friend', *yɔ́mɛɛ* 'wind') because a small number of them ending in *CV* (e.g., *róópta* 'rain', *tééta* (SG)/*túúka* (PL) 'cow') clearly still look like long forms in other Kalenjin languages; before this change, Akie short forms had conceivably gone out of use, and perchance had completely disappeared.[19] In fact, as described in Section 3.3.2, there are constructions in Kupsapiiny where such consonant deletion is observed: (i) two out of the three possessive constructions involving pronominal possessors, (ii) one of the demonstrative suffix constructions, and (iii) the construction with *(=)ake* (SG)/*(=)alak* (PL) '(a) certain, (an)other' (Section 3.3.2–[3]) as used with frequently used nouns, whose final consonants can be dropped in fast speech (e.g., *cii=ake* instead of *ciit=aake* in (33) and (34)).

---

[19] In fact, according to König, the Akie noun *akie* had the form of *akiek* until about 100 years ago, and its closely related variety is still called "Okiek".

4.3. Non-correspondence between Kupsapiiny long and short noun forms and their frequency

According to the Iconicity Principle (Haiman 1983, 1985), morphological or formal complexity reflects semantic complexity or rarity. Apparently, the two types of noun forms in Kupsapiiny systematically violate this principle – long noun forms have more morphological substance, but they occur more frequently and in a wider range of situations on the RH than their short counterparts. Because the Kupsapiiny noun system as a whole contradicts the pattern expected from the Iconicity Principle, it is a systematic version of "markedness reversal" (Waugh 1982) or "local markedness" (Tiersma 1982), which are reported to appear as sporadic counterexamples to the principle within a system: for example, in languages with marked singular and unmarked plural nouns, such nouns are those that more often occur in the plural than in the singular (Tiersma 1982, Dimmendaal 2000).[20]

Haspelmath (2006, 2021), who rejects an explanation of grammatical coding asymmetries in terms of form–meaning iconicity or markedness, proposes a frequency-based explanation of not only word forms (Zipf 1935) (e.g., the final-consonant loss mentioned in Section 4.2) but also grammatical coding patterns in the form of "the grammatical form–frequency correspondence hypothesis" as follows: "When two grammatical construction types that differ minimally (i.e., that form a semantic opposition) occur with significantly different frequencies, the less frequent construction tends to be overtly coded (or coded with more segments), while the more frequent construction tends to be zero-coded (or coded with fewer segments), if the coding is asymmetric" (2021: 606). Exemplifying "markedness reversal" or "local markedness" with marked singular and unmarked plural forms of certain nouns forming part of the lexicon of some languages (also, Haspelmath & Karjus 2017) and a marked third-person and an unmarked second-person imperatives in some languages, Haspelmath (2006: 44) states that "[t]his situation violates iconicity … but it is in perfect harmony with the frequency-based explanation", and is explicable in terms of frequency vs. rarity of use, instead of unmarkedness vs. markedness. Unlike an iconicity violation of this type, the frequent use of long noun forms in Kupsapiiny for almost all situations is pervasive through the noun system, and apparently defies the frequency-based explanation of forms.

Haspelmath (2021: 609, 610–611, 630) makes a distinction between cross-linguistic generalizations and language-particular analyses and emphasizes that the functional-adaptive explanation of universal coding asymmetries that he proposes concerns universal tendencies rather than language-particular grammatical regularities. When he states this, he mentions two kinds of language-particular cases that his explanation does not make a prediction about, (i) cases of symmetric

---

[20] However, marked singular and unmarked plural nouns in Kupsapiiny (in [b] of Table 1) do not show this property, and the number system of Kupsapiiny nouns violates the Iconicity Principle arbitrarily, rather than in a motivated way. Hence, this violation cannot be reduced to frequency of use, either.

coding, where two constructions in a semantic opposition are both equally coded (e.g., singular and plural markings in Modern Greek) or where neither is coded (e.g., no singular or plural marking in Mandarin Chinese), and (ii) cases of the absence of any general trend, where the asymmetric coding contrast between two constructions is language-specific to the extent that it is not observed as a trend in languages (e.g., English future-tense constructions with *will* and *going to* or *gonna*). Kupsapiiny long and short noun forms clearly show a coding asymmetry. However, synchronically, they are not in a strict semantic opposition, because of asymmetry, where long forms are used for all types of situations as long as they actually or potentially involve reference, and short forms are used for all types of indefinite situations. Furthermore, the default vs. indefiniteness contrast, which may be found only in Kalenjin languages, is not identified as a general trend in languages.

Nevertheless, diachronically, it is possible that long and short noun forms in Kupsapiiny contradict the form–frequency correspondence hypothesis – as discussed in Section 4.1, their difference might have developed from the definiteness distinction, in view of the historical development of the definite marker into the default noun marker in different languages, though this is not fully endorsed by the data on neighboring languages because only Pokot (Pökoot, Päkot) is described as making the definiteness distinction with long and short noun forms (Baroja 1989: 18–20) (Section 2.2).

As an example of the form–frequency correspondence related to definiteness, Haspelmath (2021: 615–616) mentions differential patient flagging in Hebrew, whereby indefinite patient nominals are more frequent than, but do not require the accusative preposition unlike definite patient nominals. However, in fact, it is not the case that object noun phrases are more frequently indefinite than definite – what is reported in the literature is that if a sentence contains an indefinite noun phrase, it is more likely to be object than subject (as pointed out by Hopper & Thompson 1980: 291 about Givón's 1979 argument based on his text counts).

Generally speaking, definite noun phrases are more frequent in discourse than indefinite ones across languages (e.g., Givón 2018 [1979], 2001, Haspelmath 1999, Skrzypek et al. 2021). On the other hand, there are languages that have a definite article but lack an indefinite article (98 out of 534 languages in Dryer 2013b), and therefore where definite noun phrases formed with a noun (and any modifier(s)) are longer than indefinite ones formed with the same noun (and the same modifier(s), if any). Therefore, definite and indefinite noun phrases in such languages generally seem to be a counterexample to the form–frequency correspondence hypothesis.

It is far from clear exactly why the economy-based explanation does not hold for definite and indefinite forms as well as forms developed from them, but a conceivable reason for this seems to be a discrepancy between the relation between speech participants and entities in the world at the beginning of new communication and that during communication. At an initial stage of communication, especially between strangers, there are much fewer entities in the world

that participants can identify together or have shared knowledge about than those that they cannot identify together or have no shared knowledge about; hence, at this stage, the former are rarer, and are likely to be expressed with longer noun forms (for example, with a demonstrative) than the latter, moving toward a form–frequency correspondence. However, once communication starts, participants talk more frequently of entities about which they already have shared knowledge than those of which they have no shared knowledge; accordingly, the former are more frequent (Givón 2018 [1979]), and could be expressed with shorter forms, but they are already expressed with longer noun forms (though they may be referred to with zero anaphors or pronouns: Givón 1995), moving in the other direction (the more frequent, the longer) of a form–frequency correspondence (the more frequent, the shorter). Therefore, the relations between the two types of entities are reversed in situations at the beginning of new communication and during communication; however, once forms are established in a linguistic system, it is difficult to alter them.

5. Conclusion

It has been shown that in Kupsapiiny, long noun forms are default forms, which are obligatory for the two types of definites on Dryer's RH and are also used for almost all the three types of indefinites; on the other hand, short forms are indefinite forms that occur in indefinite situations, but their mandatory use is confined to semantically nonspecific indefinite nominals involving negation and true-predicate nominals. In fact, long forms not only cover a wider range of distribution than short ones on the RH, but also occur more frequently than them even in indefinite situations.

The present study demonstrated that the major factor in the use of long forms as default forms is the speaker's assumption about the accessibility of the (type of) referent of the noun phrase to the hearer (or even to the speaker themself), which was semantically generalized from definiteness through their diachronic development. It hypothesized that the default vs. indefinite contrasts between long and short forms originates from the definiteness distinction, and that in the course of this development, context extension was brought about for long forms by associations between referents in discourse as well as subjectification.

There are other sorts of factors that affect the use of long and short forms, including properties specific to lexical items, constructions, and contexts as well as a generation difference that affect the use of long and short forms. Although some of these factors may appear to be unaccountable, most of them prove to be ascribable to the default or definite vs. indefinite difference or sound change. It continues to be extremely difficult to pinpoint the most influential factor in each individual instance, and to find out exactly how different factors came to interact with each other during the diachronic development. These remain to be investigated in future studies.

If the hypothesized scenario about their diachronic development from forms showing the definiteness contrast is correct, Kupsapiiny long and short noun forms

turn out to exhibit a form–frequency non-correspondence, presumably because of a mismatch between the relations between speech participants and entities in the world in pre-discourse situations and in discourse situations, which by conjecture pertains to definiteness in general.

Abbreviations

1: first person, 2: second person, 3: third person, AL: 'along, continuosly, iteratively', CMPL: complementizer, COM: comitative, COP: copula, DEF: definite, DM: discourse marker, D.PST: distant past, EMPH: emphatic, ENG: English, EP: epenthesis, H: 'hither', IMPRS: impersonal, INDEF: indefinite, IPFV: imperfective, LOC: locational, LUG: Luganda, LV: lengthened vowel, NEG: negative, NMLZ: nominalizer, O: object, PL: plural, POSS: possessive, PURP: purposive, REFL: reflexive, PRS: present, Q: question, QUOT: quotative, REL: relative clause marker, SG: singular, SWH: Swahili, T.PST: today past, TH: 'thither', VAFW: 'via/at/from/with'

References

Baroja, Tomás Herreros (1989) *Analytical grammar of the Pokot language (Kitapu ngala pokot nyo kikir)*. Trieste: Universitá di Trieste.

Beech, Mervyn W. H. (1911) *The Suk: Their language and folklore.* Oxford: Clarendon Press.

Chafe, Wallace L. (ed.) (1980) *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production.* Norwood, NJ: Ablex.

Christophersen, Paul (1939) *The articles: A study of their theory and use in English.* Copenhagen: Einar Munksgaard.

Creider, Chet A., and Jane Tapsubei Creider (1989) *A grammar of Nandi.* Hamburg: Helmut Buske Verlag.

Dimmendaal, Gerrit J. (2000) Number marking and noun categorization in Nilo-Saharan languages. *Anthropological Linguistics* 42: 214–261.

Dixon, R. M. W. (1994) *Ergativity.* Cambridge: Cambridge University Press.

Dryer, Matthew S. (2004) Noun phrases without nouns. *Functions of Language* 11: 43–76.

Dryer, Matthew S. (2005a) Definite articles. In: Martin Haspelmath, Matthew S. Dryer, David Gil, and Bernard Comrie (eds.) *The world atlas of language structures*, 154–57. Oxford: Oxford University Press.

Dryer, Matthew S. (2005b) Indefinite articles. In: Martin Haspelmath, Matthew S. Dryer, David Gil, and Bernard Comrie (eds.) *The world atlas of language structures*, 158–61. Oxford: Oxford University Press.

Dryer, Matthew S. (2013a) Definite articles. In: Matthew S. Dryer and Martin Haspelmath (eds.) (2013) *The world atlas of language structures online*, Chapter 37. http://wals.info/chapter/37 [accessed on April 1, 2022]. Munich: Max Planck Digital Library.

Dryer, Matthew S. (2013b) Indefinite articles. In: Matthew S. Dryer and Martin Haspelmath (eds.) (2013) *The world atlas of language structures online*, Chapter 38. http://wals.info/chapter/38 [accessed on April 1, 2022]. Munich: Max Planck Digital Library.

Dryer, Matthew S. (2014) Competing methods for uncovering linguistic diversity: The case of definite and indefinite articles (Commentary on Davis, Gillon, and Matthewson). *Language* 90: e232–e249.

Givón, Talmy (1995) *Functionalism and grammar.* Amsterdam: John Benjamins.

Givón, Talmy (2001) *Syntax, Volume I.* Amsterdam: John Benjamins.

Givón, Talmy (2018 [1979]) *On understanding grammar, Revised edition.* New York: Academic Press.

Greenberg, Joseph H. (1978) How does a language acquire gender markers? In: Joseph H. Greenberg, Charles A. Ferguson, and Edith A. Moravcsik (eds.) *Universals of human language, Vol. 3: Word structure*, 47–82. Stanford: Stanford University Press.

Greenberg, Joseph H. (1981) Nilo-Saharan movable-*k* as a stage III article (with a Penutian parallel). *Journal of African Languages and Linguistics* 3: 105–112.

Greenberg, Joseph H. (1991) The last stages of grammatical elements: Contractive and expansive desemanticization. In: Elizabeth Closs Traugott and Bernd Heine (eds.) *Approaches to grammaticalization, Volume I: Theoretical and methodological issues*, 301–314. Amsterdam: John Benjamins.

Grice, Paul (1975) Logic and conversation. In: Peter Cole and Jerry L. Morgan (eds.) *Syntax and semantics 3: Speech acts*, 41–58. New York: Academic Press.

Grice, Paul (1989) *Studies in the way of words.* Cambridge: Harvard University Press.

Haiman, John (1983) Iconic and economic motivation. *Language* 59: 781–819.

Haiman, John (1985) *Natural syntax: Iconicity and erosion.* Cambridge: Cambridge University Press.

Haspelmath, Martin (1999) Explaining article-possessor complementarity: Economic motivation in noun phrase syntax. *Language* 75: 227–243.

Haspelmath, Martin (2006) Against markedness (and what to replace it with). *Journal of Linguistics* 42: 25–70.

Haspelmath, Martin (2021) Explaining grammatical coding asymmetries: Form–frequency correspondences and predictability. *Journal of Linguistics* 57: 605–633.

Haspelmath, Martin, and Andres Karjus (2017) Explaining asymmetries in number marking: Singulatives, pluratives, and usage frequency. *Linguistics* 55: 1213–1235.

Hawkins, John A. (1978) *Definiteness and indefiniteness: A study in reference and grammaticality prediction.* London: Croom Helm.

Hawkins, John A. (2004) *Efficiency and complexity in grammars.* Oxford: Oxford University Press.

Heine, Bernd, and Tania Kuteva (2007) *The genesis of grammar: A reconstruction.* Oxford: Oxford University Press.

Hollis, Alfred Claud (1909) *The Nandi: Their language and folklore.* London: Oxford University Press.

Hopper, Paul J., and Sandra A. Thompson (1980) Transitivity in grammar and discourse. *Language*, 56: 251–299.

König, Christa (2006) Marked nominative in Africa. *Studies in Language* 30: 655–732.

König, Christa (2008) *Case in Africa.* Oxford: Oxford University Press.

König, Christa, Bernd Heine, and Karsten Legère (2015) *The Akie language of Tanzania: A sketch of Discourse Grammar.* Tokyo: Research Institute for Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.

König, Christa, Bernd Heine, Karsten Legère, and Ingo Heine (2020) *The Akie language of Tanzania: Texts and dictionary.* Cologne: Köppe.

Kupsapiiny Language Board and Kupsapiiny Language Development and Preservation Foundation (2017 [2011]) *An orthography and spelling guide for Kupsapiiny, Revised version.* Entebbe, Uganda: SIL International.

Larsen, Iver (1991) Sabaot noun classification. In: Franz Rottland and Lucia N. Omondi (eds.) *Nilo-Saharan, Volume 6: Proceedings of the third Nilo-Saharan Linguistics Col-*

loquium, Kisumu, Kenya, August 4–9, 1986, 143–163. Hamburg: Helmut Buske Verlag.

Lyons, John (1982) Deixis and subjectivity: *Loquor, ergo sum?* In: Robert Jarvella & Wolfgang Klein (eds.) *Speech, place, and action: Studies in deixis and related topics*, 101–124. New York: John Wiley.

Montgomery, Christine Anne (1966) The morphology of Sebei. Unpublished doctoral dissertation, University of California, Los Angeles.

Narrog, Heiko, and Bernd Heine (2021) *Grammaticalization.* Oxford: Oxford University Press.

O'Brien, Richard J., and Wim A. M. Cuyers (1975) A descriptive sketch of the grammar of Sebei. *Georgetown University Working Papers on Languages and Linguistics*, 9, 1–108. Washington, D.C.: Georgetown University Press.

Rottland, Franz (1981) The segmental morphology of Proto-Southern Nilotic. In: Lionel Bender and Thilo C. Schadeberg (eds.) *Nilo-Saharan proceedings: Proceedings of the first Nilo-Saharan Linguistics Conference, Leiden, The Netherlands, September 8–10, 1980*, 5–17. Berlin: Mouton de Gruyter.

Schuh, Russell G. (1983) The evolution of determiners in Chadic. In: Ekkehard Wolff and Hilke Meyer-Bahlburg (eds.) *Studies in Chadic and Afroasiatic linguistics*, 157–210. Hamburg: Helmut Buske.

Skrzypek, Dominika, Alicja Piotrowska, and Rafał Jaworski (2021) *The diachrony of definiteness in North Germanic.* Leiden: Brill.

Tiersma, Peter (1982) Local and general markedness. *Language* 58: 832–849.

Toweett, Taaitta (1979) *A study of Kalenjin linguistics.* Nairobi: Kenya Literature Bureau.

Traugott, Elizabeth Closs (1989) On the rise of epistemic meanings in English: An example of subjectification in semantic change. *Language* 65: 31–55.

Tucker, Archibald Norman, and Margaret Arminel Bryan (1962) Noun classification in Kalenjin: Päkot. *African Language Studies* 3: 137–181.

Tucker, Archibald Norman, and Margaret Arminel Bryan (1964) Noun classification in Kalenjin: Nandi-Kipsigis. *African Language Studies* 5: 192–247.

Tucker, Archibald Norman, and Margaret Arminel Bryan (eds.) (1966) *Linguistic analysis: The Non-Bantu languages of north-eastern Africa.* Oxford: Oxford University Press.

Waugh, Linda R. (1982) Marked and unmarked: A choice between unequals in semiotic structure. *Semiotica* 38: 299–318.

Yap, Foong Ha, Karen Grunow-Hårsta, and Janick Wrona (eds.) (2011) *Nominalization in Asian languages: Diachronic and typological perspectives.* Amsterdam: John Benjamins.

Zipf, George K. (1935) *The psycho-biology of language.* Boston, MA: Houghton Mifflin.

Zwarts, Joost (2003) The phonology of Endo: A Southern Nilotic language of Kenya. Utrecht, Netherlands: Utrecht Institute of Linguistics.

Author's contact information:
Faculty of Business and Commerce, Keio University
kawachi@keio.jp

【要　旨】

クプサビニィ語の名詞のディフォルトの長形と不定の短形
——共時的用法と通時的発展——

河内　一博

慶應義塾大学

　クプサビニィ語（南ナイル，ウガンダ）のほとんどの名詞は，単数形と複数形のそれぞれに，接尾辞の有無により，長形と短形の 2 形式がある。本研究は，名詞の長形と短形が談話でどのように使われるかを Dryer（2014）の「指示階層」を使って分析する。長形は定の領域に使われるだけでなく，ディフォルトとして，不定の領域の多くにおいても短形よりも頻繁に用いられ，短形を使わなければならない状況は指示が起こり得ないような不定の場合に限られる。長形の意味は，かつて定性であったのが，名詞句の指示対象（のタイプ）について聞き手と知識を共有していると話者が持っている想定に基づくものへと一般化され，使用の文脈が広がったという仮説を立て，この変化について意味と語用の点から説明を試みる。長形と短形の選択のその他の要因（語彙，構文，文脈，話者の世代等）も記述する。また，一般に言語に見られる傾向に反し，より複雑な長形が短形よりも頻繁に起こることも指摘する。