Experimental evidence for semantic and morphophonological productivity in Kĩĩtharaka noun classes

Patrick Kanampiu, Centre for Language Evolution, University of Edinburgh, Tharaka University, patrick.kanampiu@tharaka.ac.ke

Alexander Martin, Center for Language and Cognition, University of Groningen, alexander.martin@rug.nl

Jennifer Culbertson, Centre for Language Evolution, University of Edinburgh, jennifer.culbertson@ed.ac.uk

In nominal classification systems, both form (i.e., morphology, phonology or both) and meaning often interact to determine the class or gender of the noun. In Bantu languages in particular, linguistic analysis has often put the emphasis on meaning, both inherent and evaluative (e.g., diminutive). However, recent quantitative studies have argued that both meaning and morphophonology – the ubiquitous nominal prefixes – serve as cues to class in Bantu, with their robustness and specific aspects potentially differing across individual languages. Here, we conducted an experimental study aimed at establishing whether speakers of Kĩĩtharaka (Bantu, E54) are sensitive to both semantics and morphophonology when classifying novel Kĩĩtharaka nouns. We used two wug-task-style experiments to establish whether particular aspects of meaning or form (here, nominal prefixes) would influence participants' production of agreement on nominal dependents. Results showed that speakers are sensitive to two inherent features, *Human* and *Fruit*, and evaluative features like *Augmentative*, *Pejorative* and *Diminutive*. On the other hand, they are robustly sensitive to all morphophonological features tested when classifying novel nouns. Our results suggest that semantic features are generally less productive than morphophonology in the Kĩĩtharaka nominal classification system.

1. Introduction

Grammatical gender and noun class systems categorise nouns into classes, reflected by agreement on dependents words, such as demonstratives, adjectives and numerals (see, e.g., Aikhenvald, 2006; Corbett, 1991; Hockett, 1958; Katamba & Stonham, 2006, amongst others). In these systems, nouns that trigger the same agreement patterns are sometimes referred to as sharing agreement class, or gender. Nouns that themselves share morphophonological properties are in the same form or deriflection class (Güldemann & Fiedler, 2019).[1] In some languages, these largely overlap, but in others, they can diverge. In such cases, not all nouns which share some morphophonological feature (e.g., a prefix or suffix) will necessarily have the same gender. For example, in French, *most* nouns that end in *–ette* trigger feminine agreement on their dependents, but there are some (e.g., *squelette* 'skeleton') which trigger masculine agreement. In Bantu languages where gender is, in principle, determined by singular-plural agreement class pairs, most nouns are marked with a nominal prefix. Those which share a prefix are in the same form class, but as in French, not necessarily the same gender. **Table 1** below provides a summary description of the noun classification system of the Bantu language Kîîtharaka. This table shows, for example, that some nouns with the prefix *mû-* are in agreement classes 1/2 (gender A), but others with the same *mû-* prefix (hence, same form class), are in classes 3/4 (gender B).[2]

In addition, there are exceptional cases when the nominal prefix is not the one typically expected given the class marked by the agreement prefix. For example, in (1) and (2), a noun with a prefix that is in form class 5 and 9, respectively, triggers agreement class 1.

(1)   î-the   û-mwe
         5-father   1-one
         'one father'

(2)   n-dagitaarî   û-mwe
         9-doctor   1-one
         'one doctor'

---

[1] More explicitly, agreement classes refer to groupings of nouns based on a specific agreeing form they trigger; gender refers to nouns that share agreement classes across these–for example, nouns that trigger the same agreement in both the singular and the plural. By contrast, form classes refer to groupings of nouns based on specific nominal forms; deriflection refer to nouns that share form classes across these–for example, nouns that have the same affixes in both the singular and plural.

[2] By tradition, agreement classes in Bantu are numbered – odd numbers for singular and even numbers for plural classes. The labelling of genders (i.e., agreement class pairs) by alphabetic letters follows Carstens (1991). All Kîîtharaka examples provided here are either generated by the first author (who is a native speaker of Kîîtharaka) or come from the Kîîtharaka Bible corpus (B.T.L, 2019).

Table 1: Kîîtharaka genders, according to Kanampiu et al. (To appear), based on singular, plural agreement class pairs. Traditional classes (e.g., from Bantu literature) are included: however, there is not a perfect correspondence between form class and agreement class. This table therefore describes general trends. Note:(i) The class 1 and 3 singular agreement prefixes are almost entirely homophonous, (ii) The segments in brackets are optionally licensed by Kîîtharaka phonotactics, (iii) See footnote 4 for a comment about classes 14/15.

Gender	Traditional class pairs	Form class prefixes	Agreement class prefixes	Examples	Glosses
A	1	mû-	û-	mu-ntû û-mwe	one person
	2	a-	ba-	a-ntû ba-îrî	two people
B	3	mû-	û-	mû-tî û-mwe	one tree
	4	mî-	(y)î-	mî-tî (y)î-îrî	two trees
C	5	î-	rî-	î-guna rî-mwe	one baboon
	6	ma-	ma-	ma-gûna ma-îrî	two baboons
D	7	k(g)î-	k(g)î-	gî-kaabû kî-mwe	one basket
	8	i-	bi-	i-kaabû bi-îrî	two baskets
E	9	n-	î-	n-gûkû î-mwe	one chicken
	10	n-	(c)i-	n-gûkû ci-îrî	two chicken
F	11	rû-	rû-	rû-rigi rû-mwe	one thread
	10	n-	(c)i-	n-digi (c)i-îrî	two threads
G	12	k(g)a-	k(g)a-	ka-ana ka-mwe	one child
	13	tû-	tû-	tû-ana tû-îrî	two children
H	16	ba-	a-	ba-ntû a-mwe	one place
	17	k(g)û-	k(g)û-	gû-ntû kû-îrî	two places

At the same time, as in all noun class systems (Corbett, 1991), there is evidence in Kîîtharaka that at least some nouns in a given gender category share aspects of meaning. For instance, most human nouns are in gender A (see **Table 1** for an example). As with morphophonology, though, there are exceptions (e.g., see (3)).

(3)   kî-roria   kî-mwe
       7-prophet 7-one
       'one prophet'

This lack of perfect correspondence between the form of the noun, or its meaning, and the agreement it triggers highlights a longstanding question in research on gender and noun class systems: what exactly determines the agreement class of a noun? That is, what productive rules

do speakers of a language form regarding the features of a noun that might predict its gender or class? While it is often argued that gender is stored as a lexical feature of the noun (e.g., see V. M. Carstens, 1991, for claims about Bantu in particular), there is, nevertheless, robust evidence that learners and speakers make use of such cues productively, e.g., to determine the gender of novel nouns (or loan words) (e.g., Demuth, 1992; Gagliardi & Lidz, 2014; Karmiloff-Smith, 1981; Pérez-Pereira, 1991, and others.). Morphophonological features, other phonological properties of nouns *and* noun semantics have all been argued to function as potential cues to class (see e.g., Corbett, 1991; Kihm, n.d.; Kramer, 2015; Mathieu et al., 2018; Ralli, 2002, amongst others). However, establishing precisely what features productively determine the agreement a noun takes in a given language is not trivial, particularly given potentially high levels of exceptions (see, e.g., Amidu, 1997; Castagneto, 2017; Corbett, 1991; Kanampiu et al., To appear; Rice, 2006), and the possibility of a default class into which novel nouns can be classified regardless of form or meaning (e.g., see Dingemanse, 2006; Richardson, 1967).

This leaves us with fundamental theoretical and empirical questions, like how to determine which cues are productive for class assignment amidst such exceptions, and whether speakers of a language use these cues in a consistent way or not. Traditional approaches to Bantu nominal classes, for instance, have heavily borrowed from the so-called Proto-Bantu schema, an arguably semantically-centred system that characterises the noun classes in Bantu based on semantic features like *Human* or *Animate* and more abstract ones like *Paired-body parts*, *Liquid masses*, *Cohesive mass* and so on (see, for instance, **Table 2**). While some of these features may be relevant for some classes (e.g., human or animates for class 1/2), semantic regularity in most of the classes is imperfect at best – typically, there are either nouns that cut across diverse semantic domains or some that share features with nouns in different classes, e.g., *Diminutives*, *Augmentatives* and *Inanimates* in **Table 2**.

Most importantly, the semantically-based characterizations of Bantu are often not founded on robust analysis of synchronic empirical data. Rather, in many cases, these are descriptive approaches either based on (non-systematically picked) examples, or on native-speaker intuitions alone (see, e.g., Msaka, 2019; Worsley, 1954, for similar views and other examples). Recently, quantitative corpus-based approaches have contributed to a better understanding of cues to gender in individual Bantu languages. For example, corpus studies on Zulu and Sesotho have found that nouns with certain semantic features are indeed more frequent in some classes than others (see, e.g., Demuth, 2000; Ngcobo, 2010, and others). This suggests these features may be productively used by speakers of these languages as cues to class. Studies on Kîîtharaka have used quantitative data, including formal measures of morphological productivity (specifically, the Tolerance Principle, Yang, 2016) to show that semantic features appear to be productive for only a few classes (e.g., class 1/2) (e.g., Kanampiu et al., To appear) (see Section 2 for more details on the Tolerance Principle). By contrast, (morpho)phonology appears to be the most

Table 2: Proto-Bantu noun class system, adapted from Welmers (1973). Traditional class singular/plural pairs are based on nominal prefixes (i.e., deriflection or form classes) and do not always correspond to agreement classes. Each pair is traditionally associated with a meaning or group of meanings.

Traditional class pairs	Meaning
1/2	human, other animates
1a/3a	kinship terms, proper nouns
3/4	trees, plants, non-paired body parts, other inanimates
5/6	fruits, paired body parts, natural phenomena
6	liquid masses
7/8	manner
9/10	animates/inanimates
12/13	diminutives
14	abstract nouns, mass nouns
15	infinitives
16, 17, 18	locatives (near, remote, inside)
19	diminutives
20/22	augmentatives (diminutives)
21	augmentative pejoratives

relevant cue for the remaining classes in both Kîîtharaka, and other Bantu languages Demuth (see e.g., 2000, for Sesotho); Msaka (2019, for Chichewa); Kanampiu et al. (To appear, for Kîîtharaka). While corpus data helps make empirically informed predictions about which rules may (substantially) motivate the classification of nouns, an important question that remains to be answered is whether this reflects speakers' own language knowledge. In other words, it leaves open whether observation of human behaviour would align with corpus data.

A complementary approach is to use behavioral experiments with native speakers to assess production of gender agreement based on targeted semantic or morphophonological features. If speakers indeed make use of such information – i.e., the form of the noun or meaning – as cues to noun class, systematic patterns of behaviour in how they respond to given stimuli should be observed. Adults and even children have been found to use both noun internal and syntactic distributional information (i.e., agreeing dependents) in making decisions about grammatical categories. For example, participants can predict the likely category or class of the word, given evidence about its syntactic context or agreeing dependents (in both natural and artificial language experiments: see e.g., Gagliardi & Lidz, 2014; Redington et al., 1998; Reeder et al.,

2013). As mentioned above, in the absence of syntactic distributional information, speakers (both children and adults) can and do predict the noun class or gender of novel nouns from semantic or morphophonological information (again, in both natural and artificial language experiments, see, e.g., Frigo & McDonald, 1998; Gagliardi & Lidz, 2014; Karmiloff-Smith, 1981). Additionally, evidence suggests that different sources of information may not be used in the same way. Several experimental studies have found that, unlike adults, children acquiring nominal classes have a preference for phonological cues, even in the presence of more statistically reliable semantic cues (see, e.g., Bates et al., 1995; Culbertson et al., 2019; Gagliardi & Lidz, 2014; Karmiloff-Smith, 1981; Pérez-Pereira, 1991).

In Bantu, similar studies have found that children at certain ages have access to phonological restrictions characterizing the use of nominal prefixes (e.g., Demuth, 1992; Demuth & Ellis, 2010), suggesting that phonology is an early-available cue (see Demuth, 2000). For instance, in investigating acquisition of noun class prefixes in Sesotho, Demuth & Ellis (2010) found that children's early use (and omission) of noun class prefixes was determined by phonological features of nouns.

More recently, Lawyer et al. (2024) used behavioral experiments to investigate whether speakers of Kinyarwanda are sensitive to semantics or morphophonology when classifying nouns. The study used a triadic method (following Burton & Kirk, 1976) in which participants are shown triads of three nouns and asked to pick the odd one out. In each triad, two nouns shared either a semantic feature, a morphophonological feature, or both, with the remaining noun being different from the others. For instance, take the triad *iki-bwana* 'puppy', *in-gurube* 'pig' and *umu-kindo* 'palm tree'. Here, *iki-bwana* 'puppy' and *in-gurube* 'pig' share similar semantic features (i.e., they are animals and, specifically, mammals), while *umu-kindo* 'palm tree' is a plant. Likewise, in the triad *umu-kubuzo* 'broom', *umu-shushwe* 'rat', and *in-dimu* 'lemon', all bear different semantic properties, but the first two share a prefix. The assumption is that speakers should be able to pick out either or both of these differences, if indeed they pay attention to such features when making noun class decisions. The authors find a significant preference for semantics overall. In other words, most participants were successful at picking out semantically distinct words in triads, and used this kind of information even in trials where morphophonology could also be used.

Beyond the few studies mentioned above, there is relatively little work exploring how children and adult speakers of Bantu languages use semantic information to classify nouns, and no systematic experimental work testing the use of morphophonological cues in adult speakers. Here, we conduct two experiments investigating how adult speakers of Kîîtharaka make use of semantic and morphophonological features to predict the agreement class of novel nouns. Before reporting these experiments, we first briefly describe some key features of the noun class system of Kîîtharaka.

2. Previous work on Kîîtharaka noun classes

Kîîtharaka is an Eastern Central Kenya Bantu language spoken mainly by Atharaka people in Tharaka-Nithi County.[3] The language is mutually intelligible with other closely related languages in the Meru cluster of Bantu languages, though there are varying degrees of lexical and mophophonological differences (see, e.g., Kanana, 2011). According to descriptive accounts, the language has 17 noun classes (defined using the traditional class pairs, based on nominal prefixes), often argued to be based on abstract semantic features (as in BTL, 1993; waMberia, 1993). However, in a more recent analysis, Kanampiu et al. (To appear) showed that, as in many other Bantu languages, nominal prefixes do not perfectly predict agreement classes. In the terminology of Corbett (1991), this system exhibits crossed relations between form class and agreement class. Following Güldemann & Fiedler (2019), Kanampiu et al. (To appear) establish 8 form class pairs, or *deriflection* classes (i.e., groups of nouns sharing morphophonological features in the singular and plural) and 8 agreement class pairs, or *genders* (i.e., groups of nouns triggering the same singular/plural agreement).[4] In a recent study, Kanampiu et al. (To appear) applied the Tolerance Principle – a theory of morphological productivity (Yang, 2016, MS) – to a corpus of Kîîtharaka nouns to evaluate the productivity of selected semantic and morphophonological cues predicted to motivate nominal classification in Kîîtharaka. The Tolerance Principle is a theory of how and when learners form generalizations. It assumes that learners can store both productive rules and also a list of exceptional forms not determined by productive rules. To evaluate a potential rule, a learner considers whether the items in the lexicon that follow that rule meet a productivity threshold, taking into account the potential number of forms that would need to be listed as exceptional. The threshold, θN, is calculated using the formula N/lnN, where *(N)* is the number of items that follow the rule. A rule is productive if the number of exceptions *(e)* does not exceed θN. While the Tolerance Principle is primarily intended as a model of learning, it has also been applied to adult corpora to make predictions about productive rules (see, e.g., Björnsdóttir, 2023; Kodner, 2020; Yang, 2016, and others).

Kanampiu et al. (To appear) found that of twenty-three tested semantic features, only seven of them – *Human, Trees, Augmentative, Pejorative, Manner, Diminutive* and *Infinitive* – were productive. On the other hand, all morphophonological features tested were found to be productive; with the exception of the nasal prefix *n-*, all other singular and plural prefixes were productive, according to the Tolerance Principle. This study, therefore, suggests that (i) despite the existence

---

[3] Kîîtharaka is classified as E54 in Guthrie's classification (see Maho, 2003) and is coded as [thk], ISO 639-3 in Glottolog (Hammarström et al., 2023).

[4] There are additional agreement classes (class 14, 15 and all transumerals) which Kanampiu et al. (To appear) analyse as *general agreement marking*. These are classes in which nouns do not undergo *singular-plural* agreement class declension, which is, by convention, the basis of gender distinctions in noun classes. We have set these aside in the current study (see **Table 1**).

of exceptions, there are specific semantic and morphophological cues which should, in principle, be productive for speakers; however, (ii) morphophonology is likely to be a more robust cue to the agreement class (or gender) a noun belongs to than semantics.

It is worth noting that evaluative semantic features – *Diminutive*, *Augmentative*, and *Pejorative* – can function derivationally in Kîîtharaka. In other words, while some nouns are always in the same evaluative gender class (due to an inherent property, like their size or social stigma), it is also possible to use the evaluative prefixes and corresponding agreement on nouns that are typically in other classes. Such derivation usage is in order to express or highlight one of these evaluative features (for additional discussion of these classes in Bantu and other languages, see Di Garbo, 2013). We will return to this below. However, the key question we address in this article is whether the predictions of productivity reported in Kanampiu et al. (To appear), in fact, align with what speakers are actually sensitive to when making noun class decisions. In other words, we ask whether, when shown a novel object with a particular semantic feature, or a novel noun with a particular prefix, speakers provide agreeing forms in a systematic way, and in line with the predictions of Kanampiu et al. (To appear). If they produce systematic responses for a given feature, then we assume this feature is productive (see 3.4 for details on how we establish what counts as systematic). We, therefore, conducted a series of psycholinguistic experiments to test whether speakers are sensitive to the semantic and morphophonological cues investigated in this previous research.

We report two experiments with native Kîîtharaka speakers. These experiments follow previous research which makes use of artificial lexical items to test knowledge of nominal classes in adults and children (e.g., Bates et al., 1995; Berko, 1958; Frigo & McDonald, 1998; Gagliardi & Lidz, 2014; Karmiloff-Smith, 1981; Pérez-Pereira, 1991).[5] Our experiments sought to investigate whether adult speakers of Kîîtharaka are sensitive to particular kinds of cues – i.e., the morphophonological features of the noun, and the meaning of the noun – when producing syntactic agreement on adnominal modifiers. As we set out in Section 1, we take agreement as indicative of gender assignment. If speakers consistently use a particular semantic or morphophonological feature when providing nominal agreement, then novel nouns with that feature should be assigned the same agreement, hence, belonging to the same agreement class. Although assignment to a single agreement class does not necessarily indicate the gender of a noun (i.e., gender is a consistent *pairing* of agreement classes), we, nevertheless, make that simplifying assumption here.[6] On the other hand, if speakers do not consistently use a particular

---

[5] We chose not to use the triadic method described above in Lawyer et al. (2024), which is arguably less direct; it taps into speakers' similarity judgments rather than asking them to determine class.

[6] This is, in our view, a reasonable assumption, because for the most part these pairings are very consistent in Kîîtharaka; a given singular class is typically unique to a single gender, and the same is true of plural classes (see Kanampiu et al., To appear, for additional details).

feature when classifying nouns, no such regularity should be observed, or nouns will be assigned to a default gender. In the latter case, this is likely to be gender E (i.e., class 9/10 agreement). This is the agreement class associated with most loan nouns denoting non-humans, more particularly, those whose noun-initial (morpho)phonological make-up does not correspond to a possible prefix in Kîîtharaka. This kind of behavior is also found in many other Bantu languages, and we expect the same here (see, e.g., Demuth, 2000; Demuth & Weschler, 2012; Diercks, 2012; Rose & Demuth, 2006).

3. Experiment 1: Semantic cues

Experiment 1 tested whether Kîîtharaka speakers are sensitive to semantic cues depicted in images corresponding to novel nouns. A total of nine semantic features were investigated, following previous research in Kîîtharaka (Kanampiu et al., To appear) and related Bantu languages more generally. The set of features is shown in **Table 3**. These features target 6 genders in Kîîtharaka, corresponding to seven different singular agreement classes: 1, 3, 5, 7, 11, 12.[7] As noted in Section 2, these particular semantic features were investigated in a corpus of Kîîtharaka (see Kanampiu et al., To appear), and only a subset of them were predicted to be productive based on the Tolerance Principle: the evaluative features *Diminutive*, *Augmentative*, and *Pejorative*; and *Human*. Here we test whether these predictions are, in fact, borne out for Kîîtharaka speakers tasked with producing agreeing modifiers of novel nouns. Our experimental protocol received ethical approval from the Linguistics and English Language departmental Ethics Committee.

3.1 Participants

A total of 30 native speakers of Kîîtharaka participated in the study. These participants were recruited through personal contacts, and were mostly college students, aged between 19 and 40 years. All the participants had Kîîtharaka as their first language, and also spoke Kiswahili and English as their second or third language. Three participants used real Kîîtharaka words rather than the given novel stems throughout the task; hence, their data was excluded; data from the remaining 27 participants was retained for analysis. The experiment lasted approximately 40 minutes, and participants received 1500 Kenya shillings (KES) in compensation for their time and travel.

---

[7] Although 23 semantic features were investigated in the corpus study by Kanampiu et al. (To appear), here, we used only those that can be visually depicted on novel image stimuli. For example, the features, *Tree, Infinitive* and *Manner* were predicted to be productive in that corpus study, but these are hard, if not impossible, to depict. While the latter two are abstract, a novel tree is still a tree and can (potentially) resemble other well known trees, resulting in a confound in the experiment. Similarly, Kanampiu et al. (To appear) showed that the feature *Loan* was associated with class 9/10; however, in principle, all nouns used in our experiment could be considered as loan words by our participants. The choice for the initial set of semantic features was based on previous literature and native speaker intuitions.

Table 3: Semantic features investigated in Experiment 1, along with their respective expected gender and agreement class(es) in Kîîtharaka. Note: (i) As noted above, we expected that if participants did not pay attention to the depicted semantic feature, they would likely provide a gender E (class 9/10) agreement response; (ii) *Extended* and *Narrow* both refer to objects that are elongated, but the latter has an element of thinness (e.g., a pipe vs. a wire; a matchbox vs. a razor blade).

Semantic Feature	Expected gender	Agreement class pairs
Human	A	½
Extended	B	¾
Augmentative	C	5/6
Fruit	C	5/6
Artefact	D	7/8
Pejorative	D	7/8
Narrow	F	11/10
Wavy	F	11/10
Diminutive	G	12/13

3.2 Materials

We used novel nouns paired with novel images, in a task similar to the classic wug test paradigm (Berko, 1958). There were 27 images total, three for each of the nine target semantic features.[8]

Several notes are in order about the images and the features they test.

First, it is worth noting here that some of the features we test can be interpreted in more than one way. For example, something *Narrow* could also be *Extended* (see, e.g., **Figure 1**). Similarly, a given picture necessarily invokes more than one semantic feature. We attempted to increase the likelihood that specific images were associated with specific target features by piloting the stimuli images (for further discussion of this issue, see 3.5).

Second, images depicting evaluative features included a standard juxtaposed with an altered version of the same image (see, e.g., Sagna, 2012, for a similar manipulation of the objects to depict size). A red arrow pointed to the altered version corresponding to the novel noun label, as shown in **Figure 2**. For example, to depict *Augmentative*, two images of a novel object were next to each other, one larger than the other, and an arrow pointed to the larger one. To depict *Diminutive*, two novel objects of different sizes appeared, as in the *Augmentative*, but in this case, the arrow pointed at the smaller image. To depict *Pejorative*, a novel object was shown next to a

---

[8] We originally included three images that did not instantiate any of the target features. However, on reflection, we believe that these images were ambiguous, and probably could have been construed in many ways by participants; therefore, we do not include them here.

Figure 1: An example trial depicting the *Narrow* feature. Participants were required to type the completed phrase, and then press *îthi mbere* 'proceed' to move to the next trial.
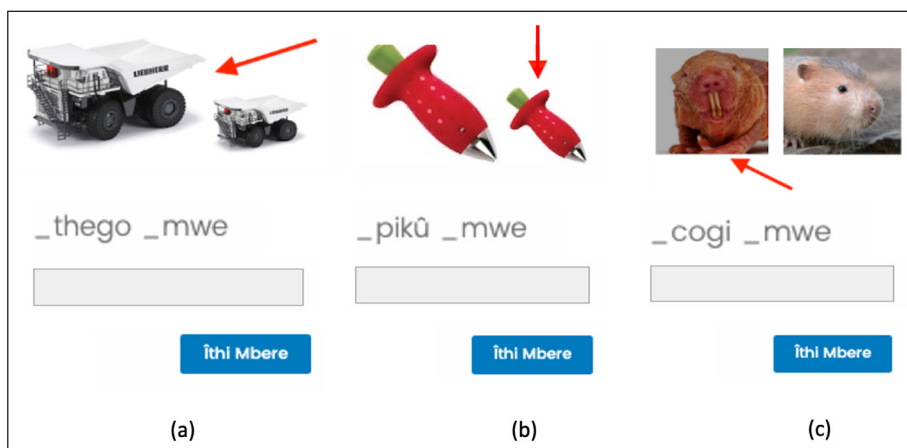

Figure 2: Example test trials depicting evaluative features for (a) *Augmentative,* (b) *Diminutive,* and (c) *Pejorative.*

malformed or disfigured version of the same type of object, with an arrow pointing at the latter. As mentioned above, evaluative features in Kîîtharaka are productive for nouns that have those features inherently, but can also be used derivationally. We opted to depict evaluative features in a way that is more similar to derivational usage (i.e., the evaluative meaning was depicted in relation to the neutral meaning), due to the difficulty of ensuring that novel nouns in isolation would convey the desired properties. For example, inherent size is difficult to convey without additional context, and pejorative meanings are, in fact, often socially determined, or based on familiarity with normative concepts.[9]

---

[9] During piloting of this experiment, we attempted to present objects conveying these meanings in isolation. Results of piloting suggested this would indeed not work well.

For the feature *Human,* we used novel (unfamiliar) human professions. Based on piloting, in these trials we included a brief description of the profession, in order to make clear to participants that this was what the novel noun corresponded to (rather than, e.g., a particular feature of the dress or costume). An example human trial is shown in **Figure 3** below. The complete set of novel object stimuli is provided in Appendix A (**Figure 10**).

Ngũgĩ: Kũringa mũtũũrirũ wa Gĩĩskoti

Ĩthi Mbere

Figure 3: An example test trial depicting human profession. The profession is indicated in red below the image as 'one who plays a Scottish flute'.

The linguistic stimuli were two-word phrases including a novel noun stem and the Kîîtharaka numeral meaning 'one'. Both appeared preceded by an underscore to indicate a missing prefix. As described below, participants' task was to provide these prefixes, and thus indicate what class they took an item to be in. As noted above, we treat the agreement prefixes as the main indicators of class. Nominal prefixes are absent, because they would otherwise provide another potential cue to class (tested in Experiment 2). However, it is rare for a noun in Kîîtharaka not to have a nominal prefix; therefore, we asked participants to provide this in their response, as well. Importantly, without either prefix, the linguistic stimuli provide no cues to class. The only hint that speakers have available is the semantic cue provided by the corresponding image.[10] The numeral 'one' was chosen to ensure that the respondents always gave singular prefixes, as number and noun class are confounded in Kîîtharaka (see **Table 1**).

There were 30 novel noun stems (see **Table 4**). All stems were bisyllabic nonce words that conformed to Kîîtharaka phonotactics. Stems were randomly paired with an image by-participant.

---

[10] It is worth noting that we are assuming here that the stem itself does not have any phonological features that are relevant for agreement (following Bryan, 2017; V. M. Carstens, 1991; Katamba et al., 2003; Nurse & Philippson, 2006; waMberia, 1993, and others).

Table 4: A list of novel noun stems used as novel Kîîtharaka nouns in Experiments 1 and 2.

Novel noun stems					
-tondi	-timû	-tiinki	-pori	-tago	-pûrû
-tebe	-thiiri	-kiithû	-pikû	-coki	-teerû
-pengi	-kithe	-kirû	-ciira	-kitho	-tuubo
-thindû	-tiirû	-ciibo	-thego	-theo	-kage
-kori	-tori	-taagi	-taki	-pegi	-coi

3.3 Procedure

This experiment was built and administered using Qualtrics online survey software. Participants received a link to the online experiment by email and accessed it using their mobile phones. Before taking part in the experiment, participants were gathered in a spacious hall in groups of ten and given verbal instructions.[11] Each participant was seated at a desk, spaced apart from other participants. Before starting the study, participants read an information sheet and signed an informed consent form agreeing to voluntarily participate in the study.

Participants were instructed that they would see images of humans and their profession (a description of which was indicated below every human image), animals or various objects or things. They were told that a Kîîtharaka speaker who was visiting a foreign country took pictures of these different things, wrote down their Kîîtharaka names and shared these with the investigators. They were told that some of the names and things would be familiar to them, but others would be new. They were also told that when the investigators looked at the names of the pictures, they thought in some cases, some piece(s) of the words were missing. The respondents were asked to use their knowledge of Kîîtharaka to judge whether there were any piece(s) missing and type in the intended full description. They were informed that they would first do some practice, using known Kîîtharaka words and familiar images.

The experiment had a training phase (presented in two blocks) and test phase, as described below. The stimuli in the training phase consisted of actual Kîîtharaka phrases (familiar nouns with the numeral 'one'), along with the corresponding familiar image. However, in the test phase, both the nominal prefix and the agreement were held out. In each of the practice blocks, each of the semantic features was represented by one trial instantiating the respective feature in a familiar image. Since we expected participants to use class 9 prefix/agreement to depict *Loan*

---

[11] We ran a fully online version, i.e., without the experimenter present, in piloting. However, participants in some cases misunderstood what they were meant to do, or had questions. We, therefore, determined that having the experimenter present was the best approach.

nouns (in case they failed to pick up the depicted feature), there were two additional trials, one with a nasal prefix and the other with a null prefix, to indicate that it was possible to use them.

In the first practice block, participants were shown descriptions in which an underscore replaced the nominal prefix only, for example, _gûkû îmwe for 'one _chicken'. They were asked to type a complete description, i.e., ngûkû îmwe. In this block, both the nasal and the null prefix (class 9) trials had an underscore placed before the (bare) stem. Placing an underscore where no prefix was needed was meant to show participants that the underscore did not necessarily indicate that anything was actually missing – that was for them to determine. Participants received feedback in the form of the expected answer on each trial. The order of trials was randomized for each participant.

In the second practice block, participants saw familiar images and two-word descriptions, just as before. Again, there was one trial for each of the nine singular classes tested. However, unlike in the first block, here the *second* word (always the Kîîtharaka word for 'one') had an underscore in place of the agreement prefix, for example, rwîgî _mwe for 'one eagle'. Participants were asked to type a complete description, i.e., rwîgî rûmwe. The order of trials was randomized for each participant. When a participant was finished with the two practice blocks, they were asked to indicate this to the experimenter (and could not proceed with the study until the experimenter entered a passcode on their device). The experimenter went around and spoke quietly with each participant to make sure they understood the task, after which they were allowed to proceed.

During the test phase, participants saw unfamiliar images, along with a description consisting of a novel noun stem and the Kîîtharaka quantifier 'one' without a prefix, both preceded by an underscore. Participants were required to complete the description, i.e., including any missing prefixes. There were 27 trials in total, three for each of the nine semantic features tested. There was a lag time (5 seconds) between when the image appeared, and when the linguistic stimuli and text box appeared in all trials. No feedback was provided. The order of trials was randomized for each participant.

Finally, there were three post-test questions which asked participants (i) to briefly explain what helped them judge when and what was missing, (ii) what languages they spoke apart from Kîîtharaka and (iii) whether Kîîtharaka was their first language. The reader is referred to 3.1 for some details related to the outcome of questions (ii) and (iii) and 3.5 for question (i).

3.4 Results

Figure 4 shows the average proportion of responses that aligned with the expected singular agreement class number for each semantic feature tested, as outlined in **Table 3**. The dotted line indicates what we consider here to be chance-level performance, calculated based on the
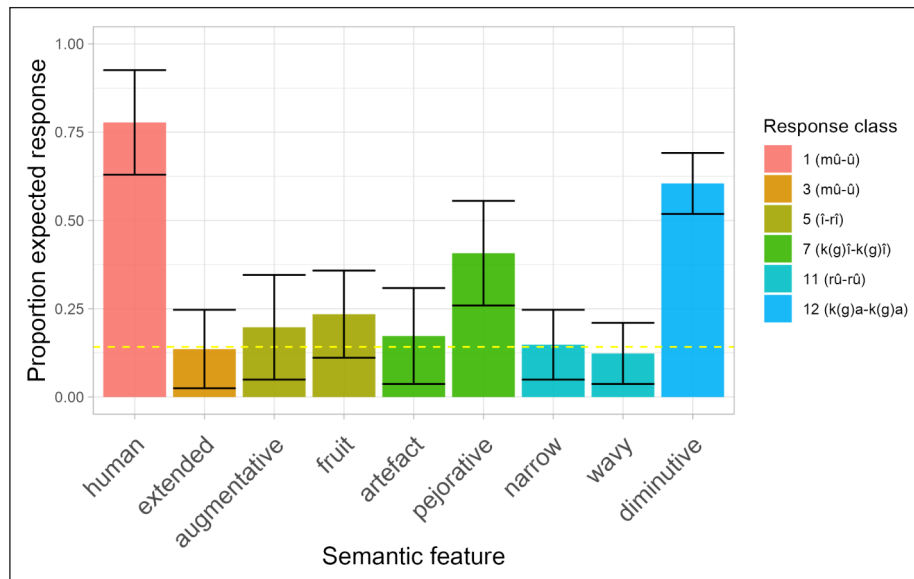
Figure 4: Average proportion of responses aligned with the predicted class (see **Table 3**) for each semantic feature. Error bars show standard error on by-participant means.

singular agreement class prefixes participants provided, which correspond to the expected classes for the 6 semantic features we tested. With 6 different expected classes, chance is $1/7 = 0.17$.[12] **Figure 4** suggests that some of the features tested may have been used productively, while others were not. Interestingly, it is also notable that a sizeable proportion of participants used class 7 prefixes ($k(g)î- – k(g)î-$) with augmentative trials, instead of the predicted class 5 ($î- – rî-$). Additionally, across all features, participants tended to produce a relatively high proportion of class 9 ($n/Ø- – î-$) responses (e.g., more than 40% for the *Artefact, Extended, Fruit, Narrow* and *Wavy* features). We come back to these patterns in 3.5, but see **Figure 11** in Appendix B for an overview of this variation in participant responses.

To further analyse these data, we evaluated whether accuracy for each of these semantic features was significantly above chance. Since *Accuracy* was a binary feature (correct vs incorrect for the expected and unexpected responses, respectively), we used *a binomial test,* with the probability of success set to 1/7. We consider accuracy significantly above this level, according to the binomial test, to indicate a productive feature. The results, provided in **Table 5**, show significantly above chance accuracy for the features *Human* ($p < 0.0001$), *Fruit* ($p < 0.0251$), *Pejorative* ($p < 0.0001$) and *Diminutive* ($p < 0.0001$). All the other features, including *Augmentative,* had accuracy rates that were not significantly above chance ($p > 0.05$).

---

[12] Note that this chance level could be somewhat conservative, since we expect participants to sometimes use the default class 9 agreement.

Table 5: Results of binomial tests of semantic features in Experiment 1. Upper and lower bounds for 95% confidence intervals around the estimate are also given.

Feature	Estimate	P-Value	Lower bound	Upper bound
Human	0.7778	<0.0001	0.6717	0.8627
Extended	0.1358	0.7367	0.0698	0.23
Augmentative	0.1975	0.1545	0.1173	0.3009
Fruit	0.2346	0.0251	0.1475	0.3418
Artefact	0.1728	0.4272	0.0978	0.273
Pejorative	0.4074	<0.0001	0.2995	0.5223
Narrow	0.1481	0.8737	0.079	0.2445
Wavy	0.1235	0.7510	0.0608	0.2153
Diminutive	0.6049	<0.0001	0.4901	0.7119

Even among features that were productive according to the binomial test, **Figure 4** suggests that participants were more sensitive to some of the features than others. Most obviously, the feature *Human* elicited numerically higher response rates compared to other features. To verify this in a way that also accounts for by-participant variation, we fitted a logistic mixed-effects regression model with *Accuracy* as the dependent variable and *Feature* as the predictor variable, along with by-subject and by-item random intercepts, using the *glmer* function of the *lme4* package (Bates, 2010) in R (R Core Team, 2023). The model uses treatment (dummy) contrast coding, with *Human* as the reference level. Therefore, this model compares mean accuracy on *Human* feature trials to mean accuracy for each other feature tested. The results of the model are shown in **Table 6**.

The results show a significant negative effect of all features in relation to *Human*. This confirms that the feature *Human* is indeed more productive than the other features tested. We discuss the implications of these results below.

Additionally, it is possible that frequency plays a role in determining participants' sensitivity to these cues. If learners more often come across nouns representing certain classes than others, they may be more likely to recruit novel nouns into those classes (see e.g., Gagliardi & Lidz, 2014, for a similar method of analysis). We, therefore, established class weights as baseline probabilities for Kîîtharaka, based on the corpus used by Kanampiu et al. (To appear), publicly available at: https://rb.gy/ows7r1. The distribution of nouns based on type frequency in this corpus is shown in **Table 7**.[13]

---

[13] We use type and not token frequency, following a wide body of literature arguing for type frequency as a fundamental variable in the productivity of grammar rules in a language (see, e.g., Aronoff, 1976; Baayen & Lieber, 1991; Hayes & Wilson, 2008; Pierrehumbert, 2001; Yang, 2016, and others).

Table 6: Results of a model comparing accuracy on the *Human* feature (reference level) to accuracy on each other feature tested.

Feature	Estimate	SE	Z-value	P-value
(Intercept)	1.3201	0.2928	4.508	<0.0001***
Extended	–3.2479	0.4351	–7.465	<0.0001***
Augmentative	–2.7832	0.4021	–6.922	<0.0001***
Fruit	–2.5752	0.3931	–6.551	<0.0001***
Artefact	–2.9655	0.4142	–7.160	<0.0001***
Pejorative	–1.7078	0.3636	–4.696	<0.0001***
Narrow	–3.1707	0.4328	–7.327	<0.0001***
Wavy	–3.3825	0.4509	–7.502	<0.0001***
Diminutive	–0.8655	0.3621	–2.390	0.0168*

Table 7: Number of nouns in each target gender, based on the Kîîtharaka corpus used by (Kanampiu et al., To appear).

Gender	Agreement Class	Noun types
A	1/2	196
B	3/4	208
C	5/6	158
D	7/8	237
E	9/10	379
F	11/10	108
G	12/13	77
Total	–	**1363**

To establish whether the probability of a correct response for each feature was higher than would be expected, on the type frequency of the target class, we conducted additional binomial tests. In this set of tests, the probability of success was calculated using the number of nouns in the target class divided by the total number of nouns for the classes tested. The results are as shown in **Table 8**.

These results largely align with the previous ones: the features *Human, Diminutive, Augmentative, Pejorative, Fruit* were found to be significantly above chance. However, here, the feature *Narrow* is also significantly above chance. We return to this difference in 3.5 below.

Table 8: Results of binomial tests of semantic features in Experiment 1, with the baseline class probabilities from the corpus as the probability of success. Upper and lower bounds for 95% confidence intervals around the estimate are also given.

Feature	Estimate	P-value	Lower bound	Upper bound
Human	0.7778	< 0.0001	0.6717	0.8627
Extended	0.1358	0.7592	0.0698	0.23
Augmentative	0.642	< 0.0001	0.1173	0.3009
Fruit	0.2346	0.0025	0.1475	0.3418
Artefact	0.1728	1.0000	0.0978	0.273
Pejorative	0.6543	< 0.0001	0.2995	0.5223
Narrow	0.1481	0.0355	0.079	0.2445
Wavy	0.1235	0.1466	0.0608	0.2153
Diminutive	0.6049	< 0.0001	0.4901	0.7119

We also conducted one additional analysis aimed at establishing the level of divergence between the distribution of nouns to respective target classes in the experiment, where semantic features were manipulated, and across all nouns in the corpus data sets, regardless of their features. To do this, we calculated Jensen-Shannon (JS) divergence between these distributions, following Gagliardi & Lidz (2014).[14] This test provides an index between 0 and 1, where 0 is no divergence, and 1 is total divergence. In our case, this index provides additional information about the relationship between participants' responses to specific features in the experiment and the baseline distribution of nouns across classes, as evident from the corpus. In principle, smaller divergence scores would suggest that participants' responses are not strongly influenced by a given feature, i.e., that their responses are distributed simply according to the type frequency of classes. Higher divergence scores would, instead, suggest that participants are sensitive to this feature, i.e., their responses diverge from what would be expected, on type frequency. The JS scores for all features are shown below (see **Table 9**).

Notably, the features *Human*, *Diminutive*, *Augmentative* and *Pejorative* exhibit the highest divergence. This aligns with the previous analyses reported above, and suggests that these features are the ones which are used most robustly by participants when determining the agreement class of novel nouns.

---

[14] Note that the distributions that were compared are slightly different from those used by Gagliardi & Lidz (2014). In Gagliardi & Lidz (2014), there is an explicit baseline condition, where nouns were given without any semantic or formal features. So, there, experimental distributions were compared to these baselines. Here, instead, we use the corpus data as a baseline.

Table 9: Jensen-Shannon divergence scores reflecting the difference between the distribution of nouns across target classes for a given feature in the experiment, and the distribution of nouns across all classes in the corpus (as a baseline distribution). Higher scores suggest more divergence from the baseline, and thus more sensitivity to the feature in determining the class of participants' response.

Feature	Agreement class	JS score
Human	1	0.1835
Extended	3	0.0588
Augmentative	5	0.2559
Fruit	5	0.0772
Artefact	7	0.0568
Pejorative	7	0.1469
Narrow	11	0.0632
Wavy	11	0.0697
Diminutive	12	0.2364

3.5 Discussion

Experiment 1 was aimed at establishing whether speakers of Kîîtharaka are sensitive to particular semantic cues when producing agreement on dependents of novel nouns without any morphophonological cues (i.e., prefixless nouns). Traditional accounts of nominal classification in Kîîtharaka and Bantu more generally associate various noun classes with particular semantic properties. However, more recent quantitative studies have suggested that morphophonology plays an important role (see, e.g., Kanampiu et al., To appear; Msaka, 2019, and others), questioning to what extent semantic features are actually productive for speakers of modern Bantu languages. For instance, Kanampiu et al. (To appear) showed that of the several semantic features potentially motivating classification in Kîîtharaka, only six (*Human*, *Trees*, *Augmentative*, *Pejorative*, *Diminutive* and *Infinitive*) are predicted to be productive under the Tolerance Principle. In this study, we, therefore, tested whether Kîîtharaka speakers, in fact, show sensitivity to the same semantic cues. We picked out those cues which could be easily represented in pictures and showed Kîîtharaka-speaking participants pictures of unfamiliar objects instantiating these features along with corresponding phrases comprised of novel noun stems and quantifiers (the number 'one'), both with missing prefixal morphology. If participants use the semantic features cued by the images, then we expected them to fill in the missing agreement (and noun class prefix) in a consistent way. Our results showed that the expected agreement prefixes were produced systematically for *Human* (on average, approximately 75% of the time) and *Diminutive* (around 60%) and *Pejorative* (about 30%). Most other nouns were assigned class 9 agreement (the default noun class), regardless of the images they were paired with.

These results largely support the predictions made based on the corpus analysis in Kanampiu et al. (To appear). First, participants were found not to be sensitive to many of the semantic features tested when providing nominal agreement. This largely aligns with the corpus results: features like *Artefact*, *Narrow*, *Wavy* and *Extended* were found to be unproductive, and our experimental results are mostly consistent with this. When the participants did not produce the expected class agreement, they either provided class 9 agreement (which was expected, since this class is associated with loan words) or other semantically heterogeneous, relatively large classes (e.g., class 3, 5, and 7).[15] That participants showed high sensitivity to the human feature is perhaps not surprising, given that this feature is generally argued to be relevant for noun class systems in all Bantu languages. However, Kanampiu et al. (To appear), in fact, suggested that the feature is only productive once human nouns with pejorative connotations were excluded, or when the Tolerance Principle is recursively applied to the subset of nouns within a particular form class (i.e., with a class 1/2 prefix). Here, stimuli are presented without a prefix, but the experimental stimuli targeted names of humans in their professions, which participants are likely to consider non-pejorative. Thus, our results suggest that non-pejorative *Human* is a productive feature for these speakers.

While they were not used as consistently as the *Human* feature, participants were also sensitive to the evaluative features, predicted to be productive, based on Kanampiu et al. (To appear). *Diminutive* and *Pejorative* were productive across all of our analyses. By contrast, *Augmentative* was not productive, based on our initial analysis, but follow-up analyses taking into account type frequency suggested that participants were, in fact, sensitive to this feature, as well.

Interestingly, a look at the responses for *Augmentative* trials suggests the possibility that slightly lower response accuracy may reflect some ambiguity in how this feature is expressed in Kîîtharaka. The typical pattern is to use gender C along with the class 5/6 nominal prefix *î–*, e.g., *muntû* 'a person' (class 1) > *î-muntû* 'a huge person' (class 5), and this has been treated as augmentation in the literature (see, e.g., Castagneto, 2017; Fuchs & van der Wal, 2022; Kavari & Marten, 2009; Ström, 2012; Taraldsen et al., 2018; Van de Velde, 2019; Van Wyk, 1957). But, actually, there are many cases where an alternative nominal prefix corresponding to class 7 (along with gender D agreement) is used to encode a similar meaning in Kîîtharaka, as shown in **Table 10**. Our results may, therefore, reflect the presence of two prefixes that variably encode augmentation in the language: the class 5 prefix (*î-*) and the class 7 counterpart (*kî-*). Although further research in necessary to confirm this, this may be an ongoing change in progress, not reflected in the corpus compiled in Kanampiu et al. (To appear).

---

[15] Assignment to these other semantically heterogeneous classes could, in principle, also reflect participants' sensitivity to other semantic details we did not intend to highlight. This is, in some sense, inevitable, since a given object can never instantiate only a single semantic feature. However, by-item response patterns did not reveal anything systematic for particular images.

Table 10: An illustration of variable use of class 5 and 7 nominal prefixes in encoding augmentative meaning. These examples were generated by the first author, a native speaker of Kîîtharaka, and then corroborated by 3 other native speakers.

Class 5 prefix	Class 7 prefix	Gloss
îroori	kîroori	a lorry
îragita	kîragita	a tractor
îrinya	kîrinya	a pit
îcembe	gîcembe	a jembe
îrao	kîrao	a big flower
îrimorimo	kîrimorimo	a fictional giant man-eater

In addition to the above nouns, which are already in the Kîîtharaka lexicon in their current form, the first author checked five other Kîîtharaka nouns that do not normally have these prefixes – *nkaarî* 'car', *nyoomba* 'house', *kurû* 'dog', *mûtî* 'tree' and *mbûri* 'goat' – with four other native speakers. When asked what they would call these entities if they thought they looked ugly, the first mention was always a *k(g)î*-prefixed noun across the four participants. When asked what they would call them if they thought they looked 'very big', two speakers used *k(g)î*-prefixed nouns across all the words and the other two had one word each with the *î*- prefix. However, on further probing, all of them said either of the forms was acceptable for the two evaluative meanings, respectively, depending on context.[16] Notably, the overwhelming first choice of *k(g)î*- to mark both *Augmentative* and *Pejorative* perhaps indicates that although both *k(g)î*- and *î*- are available for the two evaluative meanings, there is preference for the former in marking both meanings. Based on the these observations, we re-analysed our data, treating both *î*- (class 5) and *k(g)î*- (class 7) responses as correct responses for *Augmentative*, and class 5 responses as correct for *Pejorative*. A binomial test on this data reveals significantly above chance accuracy ($\hat{p} = 0.6420$, $p < 0.0001$, *lower bound* $= 0.5277$, *upper bound* $= 0.7455$). There is also a significant increase in accuracy for the latter ($\hat{p} = 0.6543$, $p < 0.0001$, *lower bound* $= 0.5404$, *upper bound* $= 0.7566$). To summarise, once we take into account a potential ongoing change to the language, all evaluative features, including *Augmentative*, are productive across all our analyses. While more data is needed to confirm this variation in Kîîtharaka, the preliminary evidence provided above suggests that this should be taken into account in future work on the language.

By the same token, it is worth considering whether other features should be reconsidered in light of the specific pattern of responses participants provided. For example, *Narrow* was found to be productive according to one of the analyses we conducted. The *Extended* and *Narrow* features

---

[16] As of now, we can't account for the context that licenses the use of each of these prefixes for either of the evaluative meanings. We leave this for future research.

are quite similar – both share an "elongated" aspect. It could be that participants treated our *Extended* stimuli as *Narrow* or vice versa. However, a look at the *Extended* images (see **Figure 12**) shows that class 11 (*rû-rû*) responses (associated with the feature *Narrow*) were quite rare. On the other hand, for two of our *Narrow* stimuli participants provided class 3 (textitmû-mî) (associated with the feature *Extended*) responses about 20% of the time. This could be taken as an indication that, at least, some participants treated these as *Extended*. However, we also see a similar proportion of class 3 responses across almost all other stimuli tested (including evaluative features, for example). We will discuss these class 3 responses further in Section 6 (to preview, we think they represent another type of default class response).

Finally, in addition to the *Human* feature, one additional inherent feature, *Fruit*, was found to be productive, according to most of the analyses we performed here. We discuss this feature further in Section 6 below. First, however, we turn to Experiment 2, which tests the productivity of morphophonological cues.

4. Experiment 2: Morphophonological cues

In contrast to semantic features, Kanampiu et al. (To appear) predict that all nominal prefixes are productive morphophonological cues to class in Kîîtharaka. In Experiment 2, we test these features. Specifically, we test whether Kîîtharaka speakers use nominal prefixes on a novel noun stem (i.e., a beginning segment or syllable) to predict the form of an agreeing dependent, in the absence of semantic information. We tested both singular and plural prefixes, repeated in **Table 11** below. As in Experiment 1, we use the agreement prefix provided by participants as an indicator of the class participants assign a novel noun to. If participants use noun beginnings in this way, we predict that they will consistently assign the predicted agreement pattern to nominal dependents. Our experimental protocol received ethical approval from the Linguistics and English Language departmental Ethics Committee.

4.1 Participants

A total of 30 Kîîtharaka native speakers participated in the study. None were participants in Experiment 1. The participants were recruited through personal contacts, and were mostly college students, aged between 18 and 40 years. All participants had Kîîtharaka as their first language and most spoke Kiswahili and English as their second or third language. The data for all 30 participants was included in the analyses. The experiment lasted approximately 40 minutes and participants received 1500 KES in compensation for their time and travel.

4.2 Materials

As in Experiment 1, the linguistic stimuli in Experiment 2 were two-word phrases including a novel noun and the Kîîtharaka numeral 'one' with its agreement prefix missing. The set of novel noun stems was the same as in Experiment 1. Critically, here we manipulated the Kîîtharaka prefixes

Table 11: Kîîtharaka nominal prefixes tested in Experiment 2. Their expected singular/plural class number is shown to the left. Their expected agreement prefix and sample nouns with agreeing numerals are provided to the right. Class 14 is marked with an asterisk, because it is not an actual plural class. We included it here because it (optionally) pluralises in class 6.

Class number	Nominal prefix	Agreement prefix	Sample nouns and agreement	Gloss
1	mû(u)-	û-	muntû û-mwe	one person
2	a-	(b)a-	antû ba-îrî	two people
3	mû-	û-	mûtî û-mwe	one tree
4	mî-	î-	mîtî yî-îrî	two trees
5	î-	rî-	îgûna rî-mwe	one baboon
6	ma-	ma-	magûna ma-îrî	two baboons
7	k(g)î-	k(g)î-	gîkaabû kî-mwe	one basket
8	i-	bi-	ikaabû bi-îrî	two baskets
9	n-/ø-	î-	ngûkû î-mwe	one chicken
10	n-/ø	(c)i-	ngûkû ci-îrî	two chickens
11	rû-	rû-	rûrigi rû-mwe	one thread
10	n-	i-	ndigi i-thatû	three threads
12	k(g)a-	k(g)a-	kaana ka-mwe	one child
13	tû-	tû-	twana tw-îrî	two children
*14	û-	bû-	ûcûrû bû-bû	this porridge

affixed to these nonce stems. **Table 11** shows the tested prefixes alongside their agreement classes. Participants' task was, again, to provide the missing agreement prefixes, and thus indicate what class they took an item to be in. No visual stimuli were used in this experiment, and, therefore, no semantic cues to the meaning of the novel nouns were provided.

Novel stems were randomly paired with prefixes, using nine different lists. In each list, a given stem was paired with a different one of the nine singular or plural prefixes. Participants were randomly assigned to one of the nine lists. For example, one participant would see the prefix *mû-* with the nonce stem *kithe,* and another would see the same prefix with the stem *tondi*.[17] It is also

---

[17] This was done, rather than completely random assignment of stems to prefixes for each participant, because of technical constraints in Qualtrics. However, it also allowed us to ensure that the resulting nouns conformed to homorganic nasal assimilation and Darl's law, which are active in Kîîtharaka. In homorganic nasal assimilation, the nasal sound is realised as [n] before alveolar stops and as [m] before labial stops, etc. This is, to some extent, noted in Kîîtharaka orthography. Darl's law, on the other hand, is a dissimilation process in which voiced consonants are devoiced in the environment of voiced consonants, e.g., /k/→ [ɣ]/ __[+voice] and vice versa (see, e.g., Uffmann, 2013; wa MBERIA, 2002). The class 7 agreement prefix, for example, has two allomorphs, [ke-] (orthographically *kî-*) and [e] (orthographically *gî-*).

important to note that, as can be seen in **Table 11**, the class 1 and 3 prefixes are orthographically the same. This means that in the absence of semantic cues, as in this experiment, we expect the participants to classify *mû-* prefixed nouns into one of the two classes, and we considered either to be correct.

4.2.1 Procedure

Experiment 2 was built and administered using Qualtrics. Participants received a link to the online experiment by email and accessed it using their mobile phones. Participants were gathered in a hall in groups of 10 and and given initial instructions verbally before accessing the experiment. Each participant was seated at a desk, spaced apart from other participants. Participants were instructed that they were going to learn new Kîîtharaka words. The experiment had three experimental blocks. In the first *singular-singular* block, we tested whether speakers use a singular Kîîtharaka prefix to predict singular agreement. In the second *plural-plural* block, we tested whether participants use a plural prefix to predict plural agreement. In the third *singular-plural* block, we tested whether participants use a singular prefix to predict *plural* agreement. Each block had a training and a testing phase.

Figure 5: Sample trial in Experiment 2, showing (a) training trials with native class 5 prefixed noun *îkombe* and (b) test trials with novel class 9 prefixed noun *nkiirû*.

In the training phase preceding each block of testing, participants saw an actual Kîîtharaka noun (including its prefix). Under the noun, they saw a carrier sentence starting with 'I have seen a ___'.[18] Participants were required to complete this sentence using the noun presented above and a numeral. In the *singular-singular* block, participants saw a familiar Kîîtharaka noun with a

---

[18] Note that Qualtrics presents multiple underscores as broken lines. We always used the same number of these both for the noun and the prefix, and we advised participants that the number of dashes did not vary with the number of letters needed to fill the gap. We used a carrier sentence like this, instead of just a noun and the numeral to serve as a context for language production, since there were no images.

singular prefix, and the numeral was 'one' in each trial. Similarly, for each trial in the test phase, they saw a novel noun with a singular prefix and the numeral 'one' (see **Figure 5a**).

In the *plural-plural* training block, participants saw familiar nouns with a plural prefix, and the numeral was 'three' (i.e., *_thatû*, as in **Figure 5b**).[19] Likewise, for each trial in the test phase, they saw a novel noun with a plural prefix, and the numeral was 'three'.[20]

In the *singular-plural* block, in each training trial participants saw a familiar noun with a singular prefix, but in this case, the numeral was 'three'. They, therefore, had to convert the noun to plural before using it to complete the sentence. Likewise, for each test trial, they saw a novel noun with a singular prefix and the numeral was 'three'. They had to convert the noun into plural before filling in the missing plural agreement prefix on the numeral.

Notably, the position of the the object noun phrase and the missing agreement prefix on the numeral were indicated using underscores (e.g., *Nkwona ___ _mwe* 'I have seen _one ___'). As in Experiment 1, there was a lag time (5 seconds) between when the noun appeared, and when the sentence and text box appeared in all trials. The order in which stimuli within a block appeared was randomized for each participant. Feedback was provided after each training trial, but there was no feedback in the test phase.

4.2.2 Results

Figure 6 and **Figure 7** show participants' accuracy at producing the expected agreement prefixes in the *singular-singular* and *plural-plural* blocks, respectively. In both these blocks, accuracy is, overall, very high. In the *singular-singular* block, where participants saw a novel noun with a singular Kîîtharaka prefix, they produced singular prefixes and respective agreements largely as predicted. Performance was perfect for the prefixes *kî-* (class 7), *n-* (class 9) and *rû-* (class 11). Recall that class 1 and 3 prefixes are homophonous. For this reason, the two classes were collapsed, and a response was treated as correct if it had class 1 or 3 agreement. Notably, however, an overwhelming majority of participants chose *î-* (class 3) agreement for the stimulus prefix *mû-*, and not *û-* (class 1). This suggests that participants assumed these novel nouns depicted objects rather than humans. We return to this point below. A similar pattern was also observed in the *plural-plural* block, where participants saw a novel noun with a plural Kîîtharaka prefix and had to produce plural agreement. Here, accuracy was perfect or near perfect for the prefixes *ma-* (class 6), *mi* (class 4), *n-* (class 10), and *tû-* (class 13).

---

[19] We avoided the numeral 'two', because it undergoes an epenthetical consonant insertion before class 10 nouns, obscuring the agreement prefix (see **Table 11**).

[20] Due to an error while creating the lists, there were 3 excess trials for the prefix *mî-* and 3 fewer for the prefix *ma-*. However, we do not think this error affected participants' decision on the appropriate agreement on the basis of the given prefixes across the experiment.
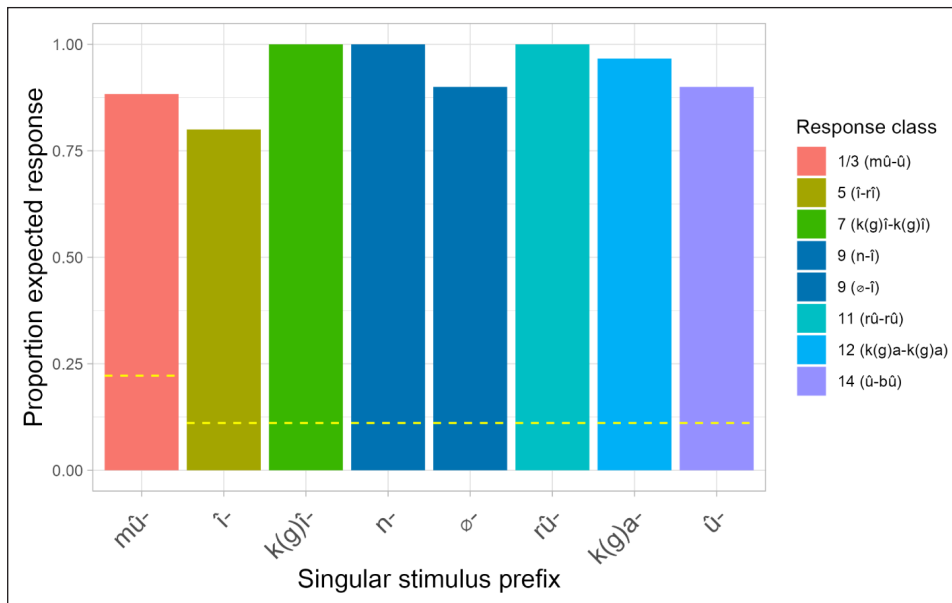
Figure 6: Average proportion of expected (correct) class agreement responses for Experiment 2 in the *singular-singular* block. Note: (i) There are no error bars, because only one prefix per class per participant was tested; (ii) Chance level is higher for the prefix *mû-* because it had, higher probability of success – participants could provide class 1/2 or 3/4 agreement (see footnote 16).
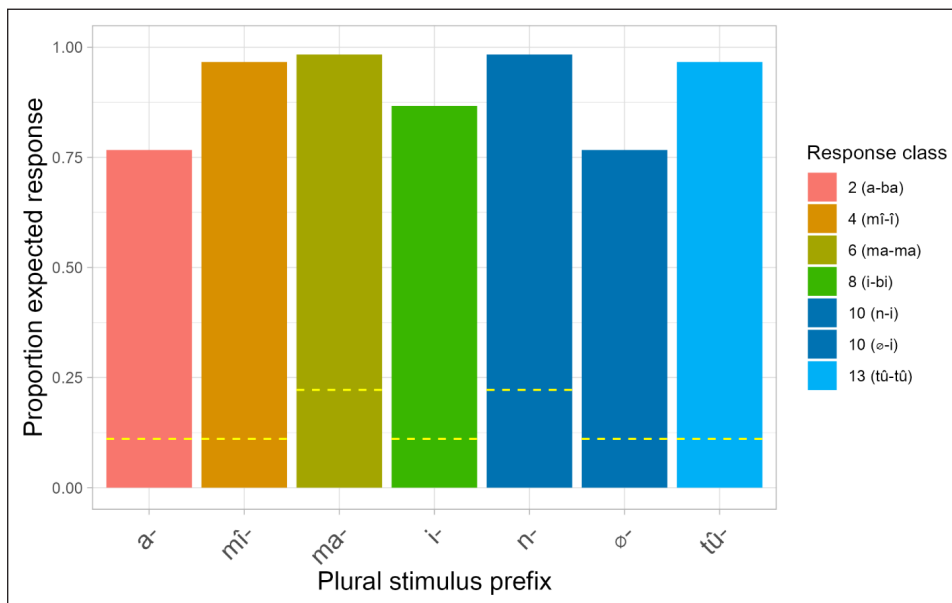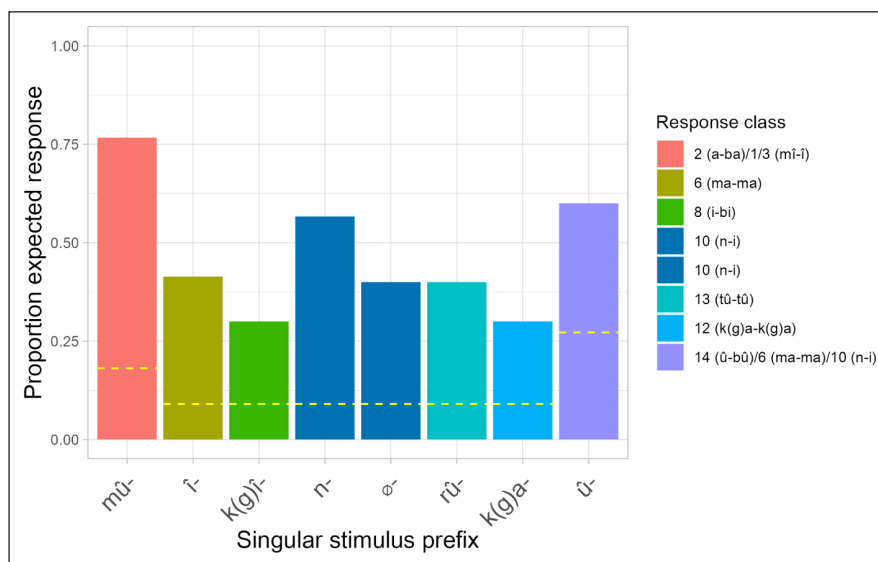

Figure 7: Average proportion of expected (correct) class agreement responses for Experiment 2 in the *plural-plural* block. Note: (i) There are no error bars because only one prefix per class per participant was tested; (ii) Chance level is higher for the prefix *ma-* and *n-*, because it had higher probability of success – participants could provide class 5/6 or 14/6 agreement for *ma-*, and 9/10 or 11/10 for *n-*.

Figure 8 shows participants' accuracy at producing the expected agreement prefixes in the *singular-plural* block. Recall that here, participants saw a singular noun and were required to convert the novel noun to plural and produce a plural agreement prefix. Accuracy rates are markedly lower for this block. In most trials where the noun had a class 1/3 prefix *mû-*, participants produced class 4 (*mî/î*) responses. In most trials where the noun had a nasal prefix, participants provided class 10 plural agreement, as expected. For several other stimulus prefixes, responses were nearly evenly divided between the expected response and class 6 agreement (e.g., the null prefix, and class 7 prefix *k(g)î-*; see **Figure 13**, **Figure 14** and **Figure 15** in Appendix B for an overview of incorrect/unexpected responses that participants gave in each of the three blocks). We return to the use of class 6 plural agreement below.

Figure 8: Average proportion of expected (correct) class agreement responses for Experiment 2 in the *singular-plural* block. Note: (i) There are no error bars, because only one prefix per class per participant was tested; (ii) Chance level is higher for the prefix *mû-* and *û-* because they had higher probability of success – participants could produce class 2 (*a-ba*) or 4 (*mî-î*) agreement for the former, and class 14 (*û-bû*), 6 (*ma-ma*) or 10 (*n-i*) for the latter.

As in Experiment 1, we evaluated whether accuracy for each of these morphophonological features (in this case, a stimulus prefix) was significantly above chance. Again, *Accuracy* was a binary feature (correct *vs* incorrect for the expected and unexpected responses, respectively), so we used *a binomial test*, with the probability of success corresponding to the probability of a correct response for the given prefix. As in Experiment 1, we consider features whose accuracy is significantly above chance level to be productive. The results, as shown in **Table 12, Table 13** and **Table 14**, indicate that accuracy for all the morphophonological features is significantly above chance level across the three blocks, hence, productive.

Table 12: Results of binomial tests for morphophonological features in the *singular–singular* block of Experiment 2. Lower and upper bounds for 95% confidence intervals around the estimate are also given.

Stimuli prefix	Estimate	P-value	Lower bound	Upper bound
mû-	0.8833	<0.0001	0.7743	0.9518
î-	0.8	<0.0001	0.6143	0.9229
k(g)î-	1	<0.0001	0.8843	1
n-	1	<0.0001	0.8843	1
ø-	0.9	<0.0001	0.7347	0.9789
rû-	1	<0.0001	0.8843	1
k(g)a-	0.9667	<0.0001	0.8278	0.9992
û-	0.9	<0.0001	0.7347	0.9789

Table 13: Results of binomial tests for morphophonological features in the *plural–plural* block of Experiment 2. Lower and upper bounds for 95% confidence intervals around the estimate are also given.

Stimuli prefix	Estimate	P-value	Lower bound	Upper bound
a-	0.7667	<0.0001	0.5772	0.9007
mî-	0.9697	<0.0001	0.8424	0.9992
ma-	0.9825	<0.0001	0.9061	0.9996
i-	0.8667	<0.0001	0.6928	0.9624
n-	0.9833	<0.0001	0.9106	0.9996
ø-	0.7667	<0.0001	0.5772	0.9007
tû-	0.9667	<0.0001	0.8278	0.9992

To test whether the participants were, overall, better in one block than the other, we fitted a logistic mixed-effects regression model with *Accuracy* as the dependent variable and *Block* as the predictor variable, along with linguistic stimuli (*ling_stimuli*) as random intercepts, using the *glmer* function of the *lme4* package.[21] The model uses treatment (dummy) contrast coding,

---

[21] In this experiment, each prefix was tested once per class per participant across the blocks, so there wasn't any by-participant variation. Prefixes that were predicted to correspond to two genders appeared once for each of the genders. These included *mû-* for gender A (classes 1/2) and gender B (3/4) in blocks 1 and 3, and *ma-* for gender C (classes 5/6) and the (optional) class 14/6 pairing in block 2.

Table 14: Results of binomial tests for morphophonological features in the *singular-plural* block of Experiment 2. Lower and upper bounds for 95% confidence intervals around the estimate are also given.

Stimuli prefix	Estimate	P-value	Lower bound	Upper bound
mû-	0.7667	<0.0001	0.6396	0.8662
î-	0.4333	<0.0001	0.2546	0.6257
k(g)î-	0.3	0.0010	0.1473	0.494
n-	0.5667	<0.0001	0.3743	0.7454
ø-	0.4	<0.0001	0.2266	0.594
rû-	0.4	<0.0001	0.2266	0.594
k(g)a-	0.3	0.0010	0.1473	0.494
û-	0.6	0.0002	0.406	0.7734

with *singular-singular* (block 1) as the reference level. Therefore, this model compares accuracy in *block 1* trials to accuracy for each of the other two blocks.[22] The results of the model are shown in **Table 15**.

Table 15: Results of a model comparing accuracy in *singular-singular* block 1 (reference level) to accuracy in *plural-plural* block 2 and *singular-plural* block 3 of Experiment 2.

Block	Estimate	SE	Z-value	P-value
(Intercept)	2.9145	0.3112	9.365	<0.0001 ***
pl-pl	–0.1709	0.3850	–0.444	0.657
Sing-pl	–2.9873	0.3700	–8.074	<0.0001 ***

The results show no significant difference between *singular-singular* and *plural-plural* blocks. However, we did observe a significant negative effect of *singular-plural* block 3 ($\beta = -2.9873$, $SE = 0.3700, z = -8.074, p < 0.0001$). This confirms that indeed, there is a significant difference between the participants' sensitivity to singular prefixes as a cue to singular class agreement and the same as a cue to plural class agreement. In the next section, we discuss the implications of these observations.

---

[22] Although this model does not allow for comparison between the plural-plural and singular-plural blocks, we do not think this causes a problem to our analyses. As can be seen in **Figures 6** and **7**, there was not much difference between participants' accuracy in the singular-singular and plural-plural blocks.

Since response accuracy in the *singular-plural* block was lower than in the other blocks, we also conducted an additional set of binomial tests, taking into account type frequency of classes (again, based on Kanampiu et al., To appear) for this block. The distribution of nouns across the respective target classes is shown in **Table 16**.[23]

Table 16: Number of nouns in each plural target gender, based on Kîîtharaka corpus used by Kanampiu et al. (To appear).

Gender	prefix	Agreement Class	Noun types
A	a-	2	188
B	mî-	4	208
C	ma-	6	143
D	i-	8	237
E	n-	10	377
E	ø	10	69
F	n-	10	107
G	tû-	13	77
Total	–	–	**1406**

The results were largely in line with the previous test, indicating significantly above-chance accuracy levels for most of the features as shown in **Table 17**.

Finally, the Jensen-Shannon (JS) divergence scores for the *singular-plural* block 3 are shown in **Table 18** below). These are all relatively high (e.g., they are similar to the JS scores for the most robustly productive semantic features in Experiment 1).

4.3 Discussion

This study aimed at establishing whether speakers of Kîîtharaka make use of noun beginnings (prefixes) as cues to agreement on the dependents of novel nouns in the absence of semantic information. To do this, we gave participants novel nouns prefixed with Kîîthraka affixes, and a Kîîtharaka sentence in which the object noun and the initial part of a numeral word (agreement) were missing. The participants were required to make use of the noun provided to fill in the sentence correctly. We looked to see whether the agreement they provided varied with the prefixes on the novel nouns in predictable ways. The results, indeed, show that the participants were sensitive to the nominal prefixes in providing the expected agreement.

---

[23] For simplicity, since the results for the two corpora were similar in our analysis of Experiment 1, here we use counts from the entire corpus only.

Table 17: Results of binomial tests of morphophonological features for *singular-plural* block 3 in Experiment 2, with the baseline class probabilities from the corpus as the probability of success. Upper and lower bounds for 95% confidence intervals around the estimate are also given. The prefix *mû-* appears in duplicate, to reflect the responses for the alternative two classes (2 and 4), both of which were correct.

Feature	Estimate	P-value	Lower bound	Upper bound
mû-	0.05	0.0572	0.0104	0.1392
mû-	0.7167	<0.0001	0.5856	0.8255
î-	0.4333	<0.0001	0.2546	0.6257
k(g)î-		0.0817	0.1473	0.494
n-	0.5667	<0.001	0.3743	0.7454
ø	0.4	<0.0001	0.2266	0.594
rû-	0.4	<0.0001	0.2266	0.594
k(g)a-	0.3	<0.0001	0.1473	0.494

Table 18: Jensen-Shannon divergence scores reflecting the difference between the distribution of nouns across various plural, in target classes (singular-plural block) in the experiment and the number of plural nouns in the corpus as baseline probability.

Feature	Agreement class	JS score
mû-	2, 4	0.2207
î-	6	0.1687
k(g)î-	8	0.1600
n-	10	0.1428
ø-	10	0.1438
rû-	10	0.1444
k(g)a-	13	0.1468

Accuracy was particularly high in blocks 1 and 2, where they used a singular noun to fill in a singular agreement prefix and a plural noun to fill in a plural agreement prefix, respectively. Results from the *singular-plural* block were not as high across the board. The highest accuracy was found for the prefix *mû*, where participants used one of the two expected responses, class 4 *mî-/î* around 75% of the time. As noted above, this choice may reflect a likely assumption that novel nouns represent objects, not humans. This can, however, not be fully confirmed from our current data, since the roles of semantics and morphophonology were tested separately. We, therefore, leave it to future research. By contrast, given the class 5 prefix *î-*, on average about

40% of responses used the expected class 6 agreement *ma-*, and for the class 7 prefix *k(g)î-*, on average about 30% of responses used the expected class 8 agreement, and about an equal proportion of responses in class 6. A similar scenario is observed for the prefixes *n-*, *k(g)a-* and *û-* (see **Figure 15** for more details).

There are a number of potential reasons for participants' relatively worse performance in the *singular-plural* block. First, it could be that this block is cognitively more taxing, since participants must convert a singular prefix to plural before determining the corresponding agreement. However, the fact that many incorrect responses involved use of class 6 plural agreement suggests that this could be a kind of default plural response, when class is unknown.[24] Default plural, especially involving the class 6 prefix *ma-*, is a common phenomenon in the Bantu language family (see, e.g., Fuchs & van der Wal, 2018; Ström, 2012; van der Wal & Fuchs, 2019, and others). A related possibility is that the participants interpreted these plurals as collective, also associated with class 6 (see, e.g., Contini-Morava, 2000; Subich et al., 2019, for Kiswahili). We leave for future study why exactly participants used this plural form instead of the regular one in this block in particular.

There are also a substantial number of cases of gender D (class 9/10) and surprisingly, gender B (class 3/4). As mentioned above, gender D (class 9/10) is associated with default classification in some Bantu languages (see, e.g., Asheli, 2015; V. Carstens, 1997; Demuth & Weschler, 2012, and others). It is less clear why gender B (class 3/4) plural agreement would be used. This gender has not been associated with default classification in the Bantu literature; however, it does manifest many of the same characteristics of a default class, similar to class 9/10. We come back to this in Section 6.

5. Comparison between semantic and morphophonological experiments

Based on the results of Experiments 1 and 2, our findings suggest a sharp difference in accuracy based on the semantic features compared to the morphophonological features tested. This is summarized in **Figure 9**, which shows average accuracy across the productive semantic features and block 1 of the morphophonological experiment.

Considering that semantics and morphophonology are the main factors which interact to determine class and or gender more generally (see e.g., Aikhenvald, 2012; Corbett, 1991; Katamba & Stonham, 2006, and others), this comparison suggests that in Kîîtharaka, morphophonology may generally be the preferred cue type. However, this claim can not be confirmed here since these two cue types were tested separately. We leave this to future research.

---

[24] Thanks to Marten Mous for sharing similar thoughts during 'This Time For Africa (TTFA)' talk at Leiden University.
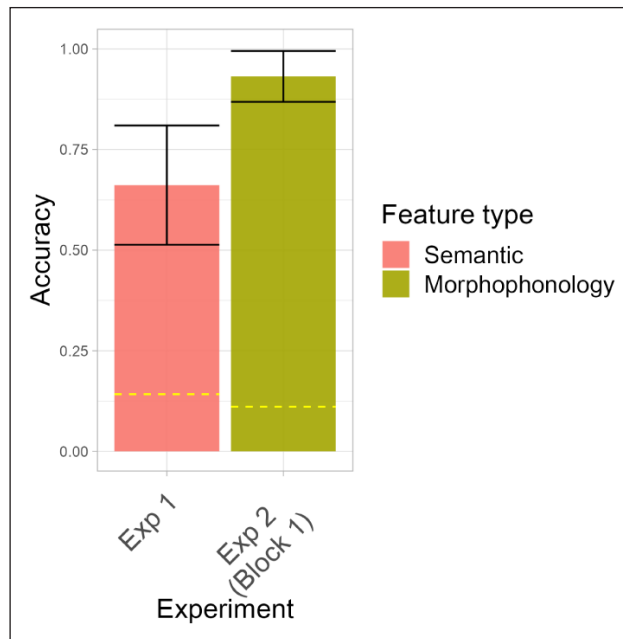
Figure 9: Average accuracy across the productive semantic and morphophonological features in Experiments 1 and 2, respectively. Chance levels vary by experiment (0.16 and 0.11, respectively).

To compare these feature types statistically, we collapsed data from the productive features of Experiment 1 (*Human, Augmentative, Pejorative,* and *Diminutive*), and from all morphophonological features in block 1.[25] We then fit a logistic mixed-effects regression model with accuracy as a binary dependent variable and experiment as the predictor variable, along with by-participant and by-item random intercepts. The predictor variable was treatment (dummy) coded, with Experiment 1 as the reference level. The model reveals a statistically significant difference between Experiment 1 and 2 ($\beta = 2.8190$, $SE = 0.3663$, $z = 7.696$, $p < 0.0001$). This confirms that participants are significantly more sensitive to morphophonology than semantics overall, even when considering only those semantic features which were used productively.

6. General discussion

In Bantu languages, both nominal prefixes and the semantics of the noun (both inherent, e.g., human, and evaluative, e.g., diminutive) are considered to determine the agreement class and therefore gender of the noun. However, recent quantitative studies in Bantu languages have shown that both of these cues may have varying degrees of relevance in particular Bantu

---

[25] We chose block 1, because the stimuli presented to participants were restricted to singular classes, just as in Experiment 1, allowing for a more direct comparison. It is worth noting again that chance levels differ for these two experiments. We are ignoring this in the comparison; however, given that chance level is higher in Experiment 1, a finding of worse performance in that experiment will be conservative.

languages (e.g., Kanampiu et al., To appear; Lawyer et al., 2024; Msaka, 2019). This study aimed to provide behavioral evidence to establish whether speakers of Kîîtharaka, an Eastern-central Bantu language, make use of both the nominal prefixes and specific semantic features of the noun in classifying novel nouns. The study involved two experiments: Experiment 1 tested whether speakers were sensitive to a set of inherent and evaluative semantic features. Experiment 2 tested for sensitivity to nominal prefixes. In both of these experiments, participants were presented with novel stimuli and required to produce class agreement on a Kîîtharaka numeral. In Experiment 1, they saw unfamiliar images highlighting one particular semantic feature accompanied by a prefix-less novel Kîîtharaka stem, along with its prefix-less numeral 'one', and were required to produce the missing prefixes. In Experiment 2, they saw a prefixed novel Kîîtharaka noun, along with its prefix-less numeral, in a carrier sentence and were required to produce the missing prefix.

The results show that some of the tested cues were consistently used to predict class agreement, while others were not. First, all nominal prefixes in Kîîtharaka were used productively by participants to predict class agreement on a nominal dependent in our experiment. This implies that in Kîîtharaka, the first segment of the noun, i.e., the nominal prefix, is a robust cue to the gender of a noun. Second, semantic features were, in general, less likely to be used productively to predict class agreement. Indeed, according to our findings, most features tested, such as *Extended*, *Artefact*, *Narrow*, etc., failed to elicit consistent agreement classes. The exceptions were *Human*, *Fruit* and the evaluative features *Augmentative*, *Pejorative*, and *Diminutive*, which were all found to be used productively by speakers (though to different degrees). As we argued in 3.5, the evaluative feature *Augmentative* is likely ambiguously expressed in Kîîtharaka, i.e., it is productively associated with two class prefixes – *î-* and *k(g)î-*. This feature was found to be robustly productive across analyses, once this pattern of usage was accounted for. Overall, then, while some semantic features are used as cues to class and gender, these features are not as robust as nominal prefixes in Kîîtharaka. Below, we discuss the implications of these results, along with a few unexpected findings.

The results of the two experiments reported here largely corroborate the findings of a corpus-based study by Kanampiu et al. (To appear), which used the Tolerance Principle to evaluate the productivity of the two types of cues in determining class and gender agreement in Kîîtharaka. While the Tolerance Principle (like other measures of productivity) can generate predictions, experimental results like these provide behavioural evidence that speakers of Kîîtharaka can and do use particular features productively. Our results suggest that many inherent semantic features are *not* represented by speakers as cues for classifying new nouns in the language. If such features were, at some point, relevant in the diachrony of the language, or the parent Proto-Bantu, most of these features have since lost that relevance. If this is true for Kîîtharaka, then it may be the case for many other present-day Bantu languages.

Interestingly, comparing the results of Kîîtharaka to those of Kinyarwanda, as in Lawyer et al. (2024), where participants were found to be more sensitive to semantics, suggests the possibility that Bantu languages may exploit these types of cues differently.[26] Studying nominal classification systems of individual languages in an empirically-grounded way is likely to reveal interesting results.

While the alignment between the corpus data and our results suggests the differences between semantics and morphophonology in Kîîtharaka are real, it is also worth revisiting a methodological consideration. It could also be the case that some of the semantic features tested appeared less productive or unproductive in the experiment because of how the participants interpreted the image stimuli. Image stimuli necessarily fail to isolate individual semantic features, in contrast to morphophonological features. Participants may, therefore, have been influenced by those features. For example, as noted in 3.5, some *Narrow* images could also be perceived as *Extended*. Second, it could be that the stimuli themselves varied in how much they instantiated the target feature. For instance, some particular images might have been more successful in depicting target meanings than others for a given feature (see **Figure 12**). For example, the target feature *Fruit,* may have been more successful at eliciting this feature compared to other features depicted by (sometimes complex) artefacts. While we did our best to select the most appropriate images (e.g., by piloting), there is likely to always be an asymmetry when comparing these two classes of features. Therefore, the possibility remains that the productivity of semantic features in the experiment is under-estimated.

We also noted some unexpected patterns in Experiment 2 that are worth additional discussion. First, we saw that participants tended to associate nouns prefixed with an ambiguous 1/2 or 3/4 class prefix with gender B (class 3/4), rather than gender A (class 1/2). This is despite the fact that in the corpus analyses (see Kanampiu et al., To appear) this prefix is productive as a cue to both gender A and B. As mentioned above, this likely indicates that in the absence of semantics, participants assume that new words refer to objects, rather than humans.

It is also worth returning to the types of responses participants consistently provided when they did *not* provide the expected responses based on class. First, the majority of such responses were class 9/10 (gender E). This is what we predicted, since this gender is typically considered to be a default, across Bantu languages. However, we also saw minimal but consistent alternative assignment to other classes, mainly class 6 (gender C plural agreement class) and gender B (class 3/4). While further research is needed to understand whether these responses indeed reflect something non-accidental, we have suggested that these may also reflect alternative

---

[26] Notably, as we highlighted in Section 1, the tasks given to participants in our experiments differed from those in Lawyer et al. (2024). It, therefore, remains possible that the differences in feature preferences in the two studies reflect this fact, rather than properties of the languages.

default classification. Loans or novel nouns that are featurally un/under-specified are typically admitted to one class – mostly class 9/10 in Bantu (see, e.g., Gunnink et al., 2015; Mous, 2019, and others). However, gender B is also a large, heterogenous class, and Kîîtharaka speakers may thus assume novel words are likely to fall into it, particularly when no semantic feature is present, or the semantic features are not productive enough to determine class. We have suggested that class 6 responses may have been used in Experiment 2, in particular, as a default plural class, since it is used to form collective plurals in Kîîtharaka. However, this, and the general trouble our participants experienced deriving plural forms from singulars in Experiment 2 deserves further study.

While we have found here that morphophonological features appear generally more productive than semantic features in Kîîtharaka, future research should also explore more explicitly their relative importance. For example, previous research has used experimental methods to show that when semantics and (morpho)phonology features are both present but provide conflicting cues, adult speakers tend to rely on semantics (e.g., Demuth & Ellis, 2010; Gagliardi & Lidz, 2014; Karmiloff-Smith, 1981; Pérez-Pereira, 1991). Whether this is the case for Kîîtharaka as well remains to be seen.

7. Conclusion

What determines productivity in noun class systems across languages is a longstanding question. While descriptive research suggests that both semantics and morphophonology are relevant, behavioral evidence of productivity in native speakers is missing for many languages. In this study, we set out to establish whether speakers of the Bantu language Kîîtharaka are sensitive to the semantics and the form of the noun (i.e., the nominal prefix) when providing agreement on the nominal dependents. We found speakers are robustly sensitive to a small subset of semantic features: *Human*, and to a lesser extent, *Fruit*, *Narrow*, and the evaluative ones *Augmentative*, *Pejorative* and *Diminutive*. By contrast, speakers were robustly sensitive to all morphophonological cues we tested. These experimental results provide psychological evidence for these cues, and suggest that if others may have been historically relevant (i.e., in Proto-Bantu), their relevance has waned for Kîîtharaka and potentially many other Bantu languages. Our study, therefore, shows that the behaviour of novel nouns can provide insights into productive morphophonological processes in a language, and thus highlights the importance of empirically grounded research in this area.

Appendix

A. Novel images used in the semantic experiment

(1) human_1   (2) human_2   (3) human_3   (4) extended_1   (5) extended_2   (6) extended_3   (7) augmentative_1

(8) augmentative_2   (9) augmentative_3   (10) fruit_1   (11) fruit_2   (12) fruit_3   (13) artefact_1

(14) artefact_2   (15) artefact_3   (16) pejorative_1   (17) pejorative_2   (18) pejorative_3

(19) narrow_1   (20) narrow_2   (21) narrow_3   (22) wavy_1   (23) wavy_2   (24) wavy_3

(25) diminutive_1   (26) diminutive_2   (27) diminutive_3

Figure 10: Novel images used in Experiment 1 with the respective features they tested.

B Cumulative bargraphs showing errors made by participants in classification of nouns in both experiments
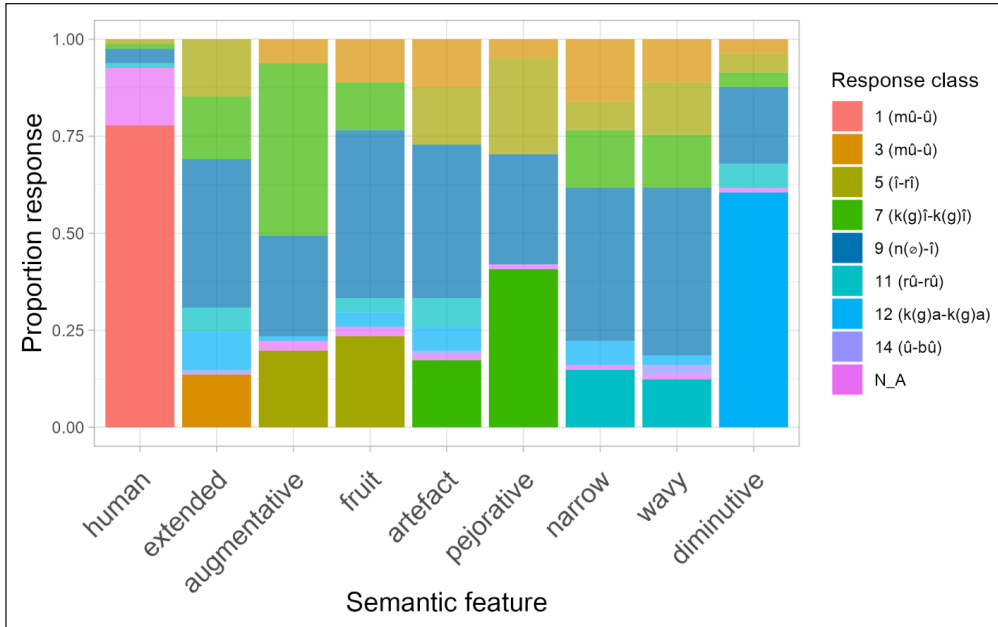
Figure 11: Cumulative responses in agreement classes by given stimuli for the semantic experiment.
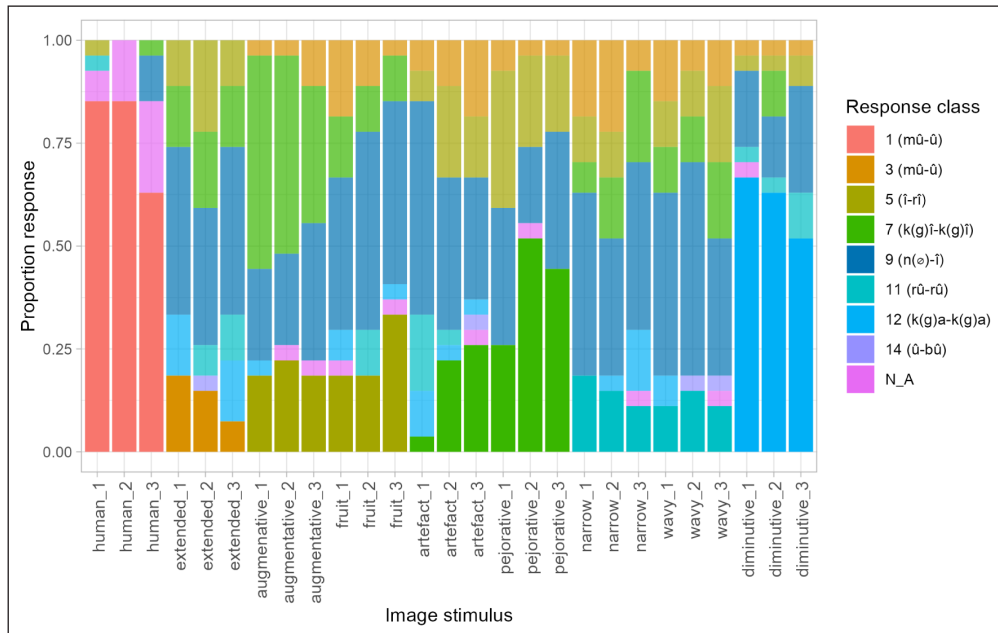
Figure 12: Cumulative responses in agreement classes by given image stimuli for the semantic experiment.
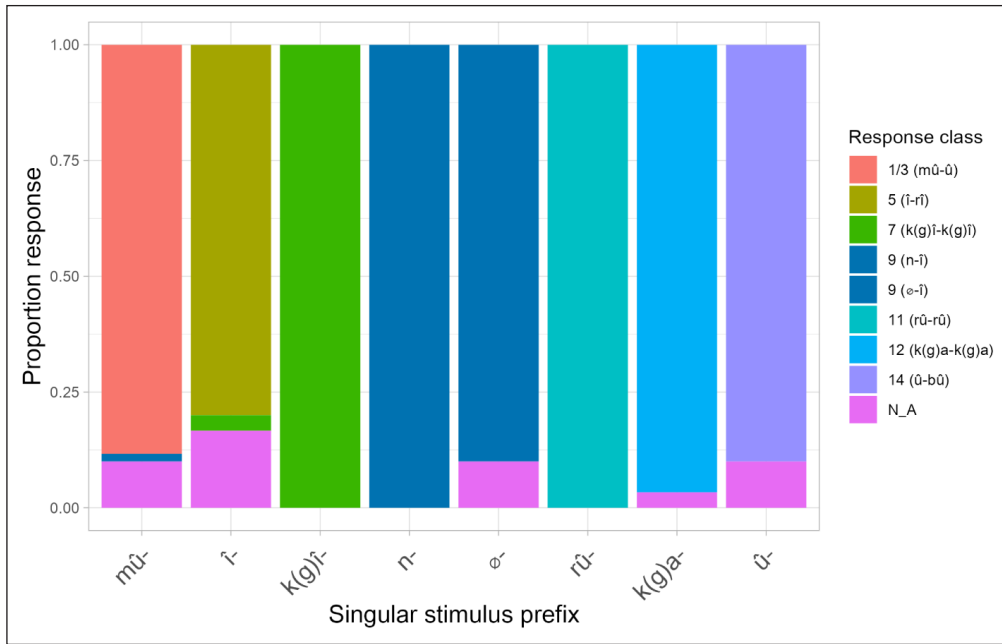
Figure 13: Cumulative responses in agreement classes by given stimuli for morphophonological experiment (*singular-singular* block).
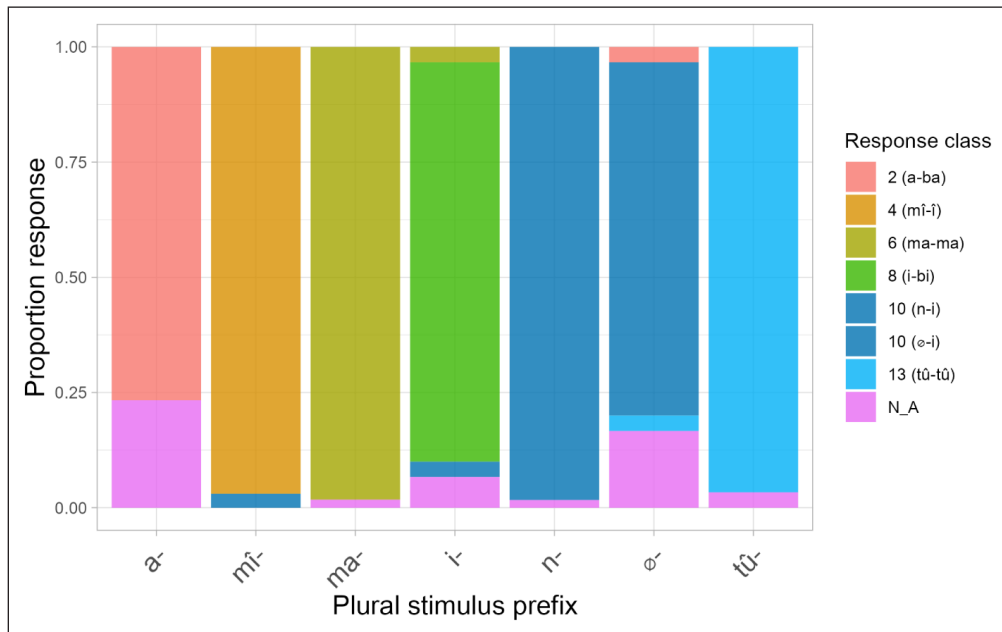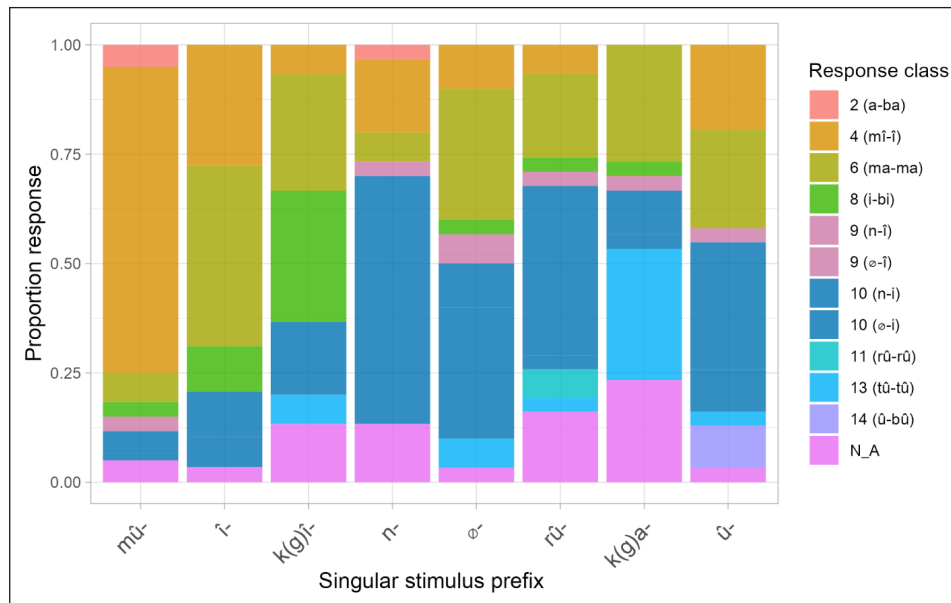

Figure 14: Cumulative responses in agreement classes by given stimuli for morphophonological experiment (*plural-plural* block).

Figure 15: Cumulative responses in agreement classes by given stimuli for morphophonological experiment (*singular-plural* block).

Data accessibility statement

All data and analyses are available online at: https://rb.gy/vcnhxs.

Ethics and consent

The ethics consent for this study was granted by the Linguistics and English Language departmental Ethics Committee (Ref. No 188-2122/4).

Acknowledgements

Many thanks to Annie Holtz for reading and providing feedback to the earlier versions of this paper.

Competing interests

The authors have no competing interests to declare.

Authors' contribution

Patrick Kanampiu: conceptualization, methodology, investigation, data collection, writing – original draft, data curation, statistical analyses, review and editing.

Alexander Martin: conceptualization, methodology, investigation, writing – review and editing, project supervision.

Jennifer Culbertson: conceptualization, methodology, investigation, statistical analyses, writing – review and editing, project supervision.

ORCIDs

Patrick Kanampiu – 0000-0002-6625-8827

Alexander Martin – 0000-0003-1241-2435

Jennifer Culbertson – 0000-0002-1737-6296

References

Aikhenvald, A. Y. (2006). Classifiers and noun classes: Semantics. In K. Brown (Ed.), *Encyclopedia of languages and linguistics* (pp. 463–471). Elsevier. https://doi.org/10.1016/B0-08-044854-2/01111-1

Aikhenvald, A. Y. (2012). Round women and long men: Shape, size, and the meanings of gender in New Guinea and beyond. *Anthropological Linguistics*, *54*(1), 33–86. https://doi.org/10.1353/anl.2012.0005

Amidu, A. A. (1997). *Classes in Kiswahili: A study of their forms and implications*. Rüdiger Köppe.

Aronoff, M. (1976). *Word formation in generative grammar*. MIT Press.

Asheli, N. (2015). Assignment of noun classes to acronyms in Kiswahili: A study of acronym-nouns in Kiswahili newspapers. *Journal of Linguistics and Language in Education*, *9*(2), 46–59.

Baayen, H., & Lieber, R. (1991). Productivity and English derivation: A corpus-based study. *Linguistics*, *29*(5), 801–843. https://doi.org/10.1515/ling.1991.29.5.801

Bates, D. M. (2010). Lme4: Mixed-effects modeling with R.

Bates, E., Devescovi, A., Pizzamiglio, L., Damico, S., & Hernandez, A. (1995). Gender and lexical access in Italian. *Perception & Psychophysics*, *57*(6), 847–862. https://doi.org/10.3758/BF03206800

Berko, J. (1958). The child's learning of English morphology. *Word, 14*(2–3), 150–177. https://doi.org/10.1080/00437956.1958.11659661

Björnsdóttir, S. M. (2023). Predicting ineffability: Grammatical gender and noun pluralization in Icelandic. *Glossa: a Journal of General Linguistics*, *8*(1). https://doi.org/10.16995/glossa.5823

Bryan, M. A. (2017). *The Bantu languages of Africa: Handbook of African languages* (Vol. 17). Routledge. https://doi.org/10.4324/9781315104959

B.T.L. (2019). *Bible in Tharaka language*. Bible Translation & Literacy (EA).

BTL, E. A. (1993). *The noun phrase in Kîîtharaka: A description of the noun class system, adjectives, demonstratives, numerals, and relative clauses in the Tharaka language of Kenya*. Bible Translation & Literacy (E.A.), Tharaka Project.

Burton, M., & Kirk, L. (1976). Semantic reality of Bantu noun classes: The Kikuyu case. *Studies in African Linguistics*, *7*(2), 157–174.

Carstens, V. (1997). Empty nouns in Bantu locatives. *The Linguistic Review, 14*, 361–410. https://doi.org/10.1515/tlir.1997.14.4.361

Carstens, V. M. (1991). *The morphology and syntax of determiner phrases in Kiswahili.* University of California, Los Angeles.

Castagneto, M. (2017). Noun classification in kiswahili. In M. Napoli & M. Ravetto (Eds.), *Exploring intensification: Synchronic, diachronic and cross-linguistic perspectives* (pp. 79–97). John Benjamins Publishing Company. https://doi.org/10.1075/slcs.189.05cas

Contini-Morava, E. (2000). Noun class as number in Swahili. In E. Contini-Morava & Y. Tobin (Eds.), *Between grammar and lexicon* (pp. 3–30). John Benjamins Publishing. https://doi.org/10.1075/cilt.183.05con

Corbett, G. (1991). *Gender.* Cambridge University Press. https://doi.org/10.1017/CBO9781139166119

Culbertson, J., Jarvinen, H., Haggarty, F., & Smith, K. (2019). Children's sensitivity to phonological and semantic cues during noun class learning: Evidence for a phonological bias. *Language, 95*(2), 268–293. https://doi.org/10.1353/lan.2019.0031

Demuth, K. (1992). The acquisition of Sesotho. In *The crosslinguistic study of language acquisition* (pp. 557–638). Psychology Press. https://doi.org/10.4324/9781315808208-11

Demuth, K. (2000). Bantu noun class systems: Loan word and acquisition evidence of semantic productivity. In G. Senft (Ed.), *Classification systems* (pp. 270–292). Cambridge University Press.

Demuth, K., & Ellis, D. (2010). Revisiting the acquisition of Sesotho noun class prefixes. In J. Guo, E. Lieven, N. Budwig, S. Ervin-Tripp, K. Nakamura, & S. Ozzcaliskan (Eds.), *Crosslinguistic approaches to the psychology of language: Research in the tradition of Dan Isaac Slobin* (pp. 93–104). Blackwell.

Demuth, K., & Weschler, S. (2012). The acquisition of Sesotho nominal agreement. *Morphology, 22*, 67–88. https://doi.org/10.1007/s11525-011-9192-7

Di Garbo, F. (2013). Evaluative morphology and noun classification: A cross-linguistic study of Africa. *SKASE Journal of Theoretical Linguistics, 10*(1).

Diercks, M. (2012). Parameterizing case: Evidence from Bantu. *Syntax, 15*(3), 253–286. https://doi.org/10.1111/j.1467-9612.2011.00165.x

Dingemanse, M. (2006). *The semantics of Bantu noun classification: A review and comparison of three approaches* [Master's thesis]. Leiden University.

Frigo, L., & McDonald, J. L. (1998). Properties of phonological markers that affect the acquisition of gender-like subclasses. *Journal of Memory and Language, 39*(2), 218–245. https://doi.org/10.1006/jmla.1998.2569

Fuchs, Z., & van der Wal, J. (2018). Nominal syntax in Bantu languages: How far can we get with gender on n?

Fuchs, Z., & van der Wal, J. (2022). The locus of parametric variation in Bantu gender and nominal derivation. *Linguistic Variation, 22*(2), 268–324. https://doi.org/10.1075/lv.20007.fuc

Gagliardi, A., & Lidz, J. (2014). Statistical insensitivity in the acquisition of Tsez noun classes. *Language*, 58–89. https://doi.org/10.1353/lan.2014.0013

Güldemann, T., & Fiedler, I. (2019). Niger-Congo noun classes conflate gender with deriflection. *Grammatical gender and linguistic complexity, 1*, 95–145. https://doi.org/10.5281/zenodo.3462762

Gunnink, H., Sands, B., Pakendorf, B., & Bostoen, K. (2015). Prehistoric language contact in the Kavango-Zambezi transfrontier area: Khoisan influence on southwestern Bantu languages. *Journal of African Languages and Linguistics, 36*(2), 193–232. https://doi.org/10.1515/jall-2015-0009

Hammarström, H., Forkel, R., Haspelmath, M., & Nordhoff, S. (2023). Glottolog 4.8. max Planck Institute for Evolutionary Anthropology. *Online v.:* http://glottolog.org

Hayes, B., & Wilson, C. (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic inquiry, 39*(3), 379–440. https://doi.org/10.1162/ling.2008.39.3.379

Hockett, C. F. (1958). *A course in modern linguistics*. Oxford & IBH. https://doi.org/10.1111/j.1467-1770.1958.tb00870.x

Kanampiu, P., Martin, A., & Culbertson, J. (To appear). Semantic and morphophonological productivity in Kîîtharaka gender system: A quantitative study. *Glossa: A Journal of General Linguistics*.

Kanana, F. E. (2011). Meru dialects: The linguistic evidence. *Nordic journal of African studies, 20*(4), 28–28. https://doi.org/10.53228/njas.v20i4.182

Karmiloff-Smith, A. (1981). *A functional approach to child language: A study of determiners and reference* (Vol. 24). Cambridge University Press.

Katamba, F., et al. (2003). Bantu nominal morphology. In N. Derek & G. Philipson (Eds.), *The Bantu languages* (pp. 103–120). Routledge.

Katamba, F., & Stonham, J. (2006). *Modern linguistics: Morphology*. Macmillan. https://doi.org/10.1007/978-1-137-11131-9

Kavari, J., & Marten, L. (2009). Multiple noun class prefixes in otjiherero [Unpublished Manuscript].

Kihm, A. (n.d.). What's in a noun: Noun classes, gender, and nounness [Manuscript].

Kodner, J. (2020). Language acquisition in the past [Doctoral dissertation].

Kramer, R. T. (2015). *The morphosyntax of gender*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199679935.001.0001

Lawyer, L. A., O'Gara, F., Ngoboka, J. P., van Boxtel, W., & Jerro, K. (2024). Meaning or morphology: Individual differences in the categorization of Kinyarwanda nouns. *Glossa Psycholinguistics, 3*(1), 1–27. https://doi.org/10.5070/G6011226

Maho, J. (2003). A classification of the Bantu languages: An update of Guthries referential system. In *The bantu languages* (pp. 639–651). Routledge London.

Mathieu, É., Dali, M., & Zareikar, G. (2018). *Gender and noun classification* (Vol. 71). Oxford University Press. https://doi.org/10.1093/oso/9780198828105.001.0001

Mous, M. (2019). Language contact. In M. Van de Velde, K. Bostoen, D. Nurse, & G. Philippson (Eds.), *The Bantu languages* (pp. 355–380). Routledge. https://doi.org/10.4324/9781315755946-12

Msaka, P. K. (2019). *Nominal classification in bantu revisited: The perspective from Chichewa* [Doctoral dissertation]. Stellenbosch University.

Ngcobo, M. N. (2010). Zulu noun classes revisited: A spoken corpus-based approach. *South African Journal of African Languages, 30*(1), 11–21. https://doi.org/10.1080/02572117.2010.10587332

Nurse, D., & Philippson, G. (2006). *The Bantu languages*. Routledge. https://doi.org/10.4324/9780203987926

Pérez-Pereira, M. (1991). The acquisition of gender: What Spanish children tell us. *Journal of child language, 18*(3), 571–590. https://doi.org/10.1017/S0305000900011259

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. *Typological studies in language, 45*, 137–158. https://doi.org/10.1075/tsl.45.08pie

R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ralli, A. (2002). The role of morphology in gender determination: Evidence from Modern Greek. *Linguistics*. https://doi.org/10.1515/ling.2002.022

Redington, M., Chater, N., & Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science, 22*(4), 425–469. https://doi.org/10.1207/s15516709cog2204_2

Reeder, P. A., Newport, E. L., & Aslin, R. N. (2013). From shared contexts to syntactic categories: The role of distributional information in learning linguistic form-classes. *Cognitive Psychology, 66*(1), 30–54. https://doi.org/10.1016/j.cogpsych.2012.09.001

Rice, C. (2006). Optimizing gender. *Lingua, 116*(9), 1394–1417. https://doi.org/10.1016/j.lingua.2004.06.013

Richardson, I. (1967). Linguistic evolution and Bantu noun class system. In *La classification nominale dans les langues negro africaines* (pp. 373–390). Editions du Centre National de la Recherche Scientifique.

Rose, Y., & Demuth, K. (2006). Vowel epenthesis in loanword adaptation: Representational and phonetic considerations. *Lingua, 116*(7), 1112–1139. https://doi.org/10.1016/j.lingua.2005.06.011

Sagna, S. (2012). Physical properties and culture-specific factors as principles of semantic categorisation of the Gújjolaay Eegimaa noun class system. *Cognitive Linguistics, 23*(1), 129–163. https://doi.org/10.1515/cog-2012-0005

Ström, E.-M. (2012). The increasing importance of animacy in the agreement systems of Ndengeleko and other Southern Coastal Bantu languages. In B. Niclas, H. Arthur, K. Anastasia, L. Håkan, & S. Jan-Olof (Eds.), *Language documentation and description* (pp. 265–285). SOAS. https://doi.org/10.25894/ldd198

Subich, V. G., Mingazova, N. G., & Al-foadi, R. A. (2019). Main tendencies in the category of number in Japanese and Swahili. *Opción: Revista de Ciencias Humanas y Sociales, 22*, 699–714.

Taraldsen, K. T., Taraldsen Medová, L., & Langa, D. (2018). Class prefixes as specifiers in Southern Bantu. *Natural Language & Linguistic Theory*, *36*, 1339–1394. https://doi.org/10.1007/s11049-017-9394-8

Uffmann, C. (2013). Set the controls for the heart of the alternation: Dahls Law in Kitharaka. *Nordlyd*, *40*(1), 323–337. https://doi.org/10.7557/12.2508

Van de Velde, M. (2019). Nominal morphology and syntax. *The Bantu languages* (pp. 237–269). https://doi.org/10.4324/9781315755946-8

van der Wal, J., & Fuchs, S. (2019). Gender on n in Bantu DP structure: From root derived nominals to locatives [Available at https://www.zuzannazfuchs.com/uploads/1/2/9/6/129696682/bantu_camcos_8.pdf].

Van Wyk, E. (1957). An augmentative noun class in Tsonga. *African Studies*, *16*(1), 25–36. https://doi.org/10.1080/00020185708707010

wa MBERIA, K. (2002). Nasal consonant processes in Kitharaka. *Nordic Journal of African Studies*, *11*(2), 11–11. https://doi.org/10.53228/njas.v11i2.352

waMberia, K. (1993). *Kitharaka segmental morphophonology with special reference to the noun and the verb* [Doctoral dissertation]. University of Nairobi.

Welmers, W. (1973). African language structures. Berkeley (usa).

Worsley, P. M. (1954). Noun-classification in Australian and Bantu: Formal or semantic? *Oceania*, *24*(4), 275–288. https://www.jstor.org/stable/40328946

Yang, C. (2016). *The price of linguistic productivity: How children learn to break the rules of language*. MIT Press. https://doi.org/10.7551/mitpress/9780262035323.001.0001

Yang, C. (MS). A users guide to the tolerance principle [Unpublished manuscript]. ling.auf.net/lingbuzz/004146