A theory of morphological productivity is essential in characterizing noun classes: Corpus and experimental evidence from Bantu

Patrick Kanampiu, Alexander Martin, Jennifer Culbertson

TTFA Talk Series, Leiden University
(April 14, 2023)

Noun class systems

- Nouns grouped based on agreement (Hockett 1958, Corbett 1991, Katamba 2006)
- Used interchangeably with gender (Corbett, 1991:1; Aikhenvald, 2006)

Kĩĩtharaka

(1) **mû**thaka **û**ra
1.young man 1.dem.dist

(2) **kĩ**banga **kĩ**ra
7.machete 7.dem.dist

(3) **i**banga **bi**ra
8.machete 8.rel

French

(4) **le** vieil homme
Det.M old.M man.M

(5) **la** camionn**ette**
Det.F van.F

Noun class systems

- Nouns grouped based on agreement (Hockett 1958, Corbett 1991, Katamba 2006)
- Used interchangeably with gender (Corbett, 1991:1; Aikhenvald, 2006)

Kĩĩtharaka

(1)	mûthaka	**û**ra
	1.young man	1.dem.dist
(2)	kîbanga	**kî**ra
	7.machete	7.dem.dist
(3)	ibanga	**bi**ra
	8.machetes	8.dem.dist

French

(4)	**le**	vieil	homme
	Det.M	old.M	man.M
(5)	**la**	camionn**ette**	
	Det.F	van.F	

Noun class systems

- Nouns grouped based on agreement (Hockett 1958, Corbett 1991, Katamba 2006)
- Used interchangeably with gender (Corbett, 1991:1; Aikhenvald, 2006)

Kĩĩtharaka

(1) **mû**thaka **û**ra
1.young man 1.dem.dist

(2) **kî**banga **kî**ra
7.machete 7.dem.dist

(3) **i**banga **bi**ra
8.machete 8.rel

French

(4) **le** vieil homme
Det.M old.M man.M

(5) **la** camionnette
Det.F van.F

What determines noun class in language?

- **Semantics**
- **Form of the noun** (morphology and/or phonology)

Kîîtharaka

(1) **mû**thaka **û**ra ← semantics (human)
1.young man 1.dem.dist

(2) **kî**banga **kî**ra
7.machete 7.dem.dist ← morpho-phonology (kî-)

(3) **i**banga **bi**ra
8.machete 8.rel

French

(4) **le** vieil homme ← semantics (male)
Det.M old.M man.M

(5) **la** camionn**ette** ← morpho-phonology (*-ette*)
Det.F van.F

What determines noun class in language?

- **Semantics**
- **Form of the noun** (morphology and/or phonology)

Kîîtharaka

(1) mûthaka **û**ra ← semantics (human)
1.young man 1.dem.dist

(2) **kî**banga **kî**ra ← morpho-phonology (*-ete*)
7.machete 7.dem.dist

(3) **i**banga **bi**ra
8.machete 8.rel

French

(4) **le** vieil homme ← semantics (male)
Det.M old.M man.M

(5) **la** camionn**ette** ← morpho-phonology (*-ete*)
Det.F van.F

What determines noun class in language?

- **Semantics**
- **Form of the noun** (morphology and/or phonology)

Kĩĩtharaka

(1) **mû**thaka **û**ra ← semantics (male)
1.young man 1.dem.dist

(2) **kî**banga **kî**ra ← morpho-phonology (*kî/kî-i/bi*)
7.machete 7.dem.dist

(3) **i**banga **bi**ra
8.machetes 8.rel

French

(4) **le** vieil homme ← semantics (male)
Det.M old.M man.M

(5) **la** camionn**ette** ← morpho-phonology (*-ette*)
Det.F van.F

What determines noun class in a language?

- But there are typically, many exceptions!...

Kîîtharaka

(6) mûthaka **û**ra ← semantics (human)
1.young man 1.dem.dist

(7) kîroria **kî**îra ← ?
7.prophet 7.dem.dist

(8) îthe **ûûra**
5.father 1.dist.dem

(9) nkoma aathi ?
9.devil 1.went

French

(10) **le** vieil homme ← semantics (male)
Det.M old.M man.M

(11) **la** personne ← ?
Det.F person.F

(12) **la** camionn**ette**
Det.F van.F

(13) **le** squel**ette** morpho-phonology (*-ette*)
Det.M skeleton.M ?

What determines noun class in a language?

- But there are typically, many exceptions!...

Kĩĩtharaka

(6) mûthaka **û**ra ← semantics (human)
1.young man 1.dem.dist

(7) kîroria **kĩĩ**ra ← ?
7.prophet 7.dem.dist

(8) **î**the **ûûra**
5.father 1.dist.dem

(9) nkoma aathi ?
9.devil 1.went

French

(10) **le** vieil homme ← semantics (male)
Det.M old.M man.M

(11) **la** personne ← ?
Det.F person.F

(12) **la** camionn**ette**
Det.F van.F

(13) **le** squel**ette**
Det.M skeleton.M

morpho-phonology (*-ette*)

?

What determines noun class in a language?

- But there are typically, many exceptions!...

Kîîtharaka

(6) **mû**thaka **û**ra ← semantics (human)
1.young man 1.dem.dist

(7) **kî**banga **kî**ra ← morpho-phonology (*kî-*)
7.machete 7.dem.dist

(8) **î**the **û**ra ← ?
5.father 1.dist.dem

(9) **n**koma **û**yû ← ?
9.devil 1.this

French

(10) **le** vieil homme ← semantics (male)
Det.M old.M man.M

(11) **la** personne ← ?
Det.F person.F

(12) **la** camionn**ette**
Det.F van.F

(13) **le** squel**ette**
Det.M skeleton.M

morpho-phonology (*-ette*)

?

What determines noun class in a language?

- But there are typically, many exceptions!...

Kîîtharaka

(6) **mû**thaka **û**ra ← semantics (human)
1.young man 1.dem.dist

(7) **kî**banga **kî**ra ← morpho-phonology (*kî-*)
7.prophet 7.dem.dist

(8) îthe **û**ra ← ?
5.father 1.dist.dem

(9) nkoma **û**yûi ← ?
9.devil 1.this

French

(10) **le** vieil homme ← semantics (male)
Det.M old.M man.M

(11) **la** personne ← ?
Det.F person.F

(12) **la** camionn**ette** ← morpho-phonology (*-ette*)
Det.F van.F

(13) **le** squel**ette** ← ?
Det.M skeleton.M

Theoretical and empirical questions…

- How do we know what semantic or morphophonological cues are productive amidst such exceptions?
- Do speakers treat different types of cues the same way? (e.g., Karmiloff-Smith, 1981; Perez-Pereira 1991; Gxilishe et al., 2009; Gagliardi & Lidz 2014 a.o.)
- Do traditional accounts of noun classification in Bantu do answer these questions?

Required:

- A **theory** of productivity
- **Empirical evidence** for speakers' use of different cues.

Theoretical and empirical questions…

- How do we know what semantic or morphophonological cues are productive amidst such exceptions?
- Do speakers treat different types of cues the same way? (e.g., Karmiloff-Smith, 1981; Perez-Pereira 1991; Gxilishe et al., 2009; Gagliardi & Lidz 2014 a.o)
- Do traditional accounts of noun classification in Bantu answer these questions?

Required:

- A **theory** of productivity
- **Empirical evidence** for speakers' use of different cues.

Traditional approaches to Bantu noun classes

The Proto-Bantu System (Richardson, 1967; Welmers, 1973)

Noun Classes	Meaning
1/2	human, other animates
1a/2a	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

Traditional approaches to Bantu noun classes

Key features:

- Classes marked by the prefix (rather than agreement) – problematic

Swahili (Carstens 1991)

(13) **ki**faru **m**dogo
7.rhino 1.small

Kĩîtharaka (corpus data)

(14) **n**koma **û**yû
9.devil 1.this

- Centred on semantics – problematic (lack of semantic regularity for some classes, several classes sharing semantic features)

- Do not make use of any theory of productivity

Traditional approaches to Bantu noun classes

Key features:

- Classes marked by the prefix (rather than agreement) – problematic

Swahili (Carstens 1991)

(13) **ki**faru **m**dogo
7.rhino 1.small

Kĩĩtharaka (corpus data)

(14) **n**koma **a**thi
9.devil 1.went

- Centred on semantics – problematic (lack of semantic regularity for some classes, several classes sharing semantic features)

- Do not make use of any theory of productivity

Traditional approaches to Bantu noun classes

The Proto-Bantu System (Richardson, 1967; Welmers, 1973)

Noun Classes	Meaning
1/2	human, other animates
1a/2a	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	augmentatives pejoratives

Traditional approaches to Bantu noun classes

Key features:

- Classes marked by the prefix (rather than agreement) – problematic

Swahili (Carstens 1991)

(13) **ki**faru **m**dogo
7.rhino 1.small

Kĩĩtharaka (corpus data)

(14) **n**koma **a**thi
9.devil 1.went

- Centred on semantics – problematic (lack of semantic regularity for some classes, several classes sharing semantic features)

- Do not make use of any theory of productivity

Theoretical and empirical questions…

- How do we know what semantic or morphophonological cues are productive amidst such exceptions?
- Do speakers treat different types of cues the same way? (e.g., Karmiloff-Smith, 1981; Perez-Pereira 1991; Gxilishe et al., 2009; Gagliardi & Lidz 2014 a.o)
- Do traditional approaches to Bantu noun classes answer these questions?

Required:

- A **theory** of productivity
- **Empirical evidence** for speakers' use of different cues.

A different approach to Kĩîtharaka

- The project aims to establish productive features that determine noun class membership in Kĩîtharaka:
  - Corpus analyses using Tolerance Principle (Yang, 2016)- provides a way of establishing productivity of rules in a set of items with exceptions (exceptions should not exceed a certain threshold)
  - Psycholinguistic tests – shows what speakers pay attention to.
  - Includes testing the relative importance of semantic and morpho-phonological cues (following e.g., Karmiloff-Smith, 1981; Gagliardi & Lidz 2014, and others).

Building a Corpus

- The study uses a corpus (2327 nouns)
  - ✓ 901 collected from Kĩĩtharaka bible (published in 2019) and,
  - ✓ 1426 translated from SIL comparative African Wordlist (Snider and Roberts, 2006).
  - Translation of SIL list done to achieve a balanced corpus
- Coded for singular/plural agreement, semantics and morphophonological features (prefixes).

Semantic features coded for

Expected Class	Semantic Feature(s)
1	human
3	Extended shape, spread shape, protruding shape, trees & plants, dispersive mass
5	Fruits, round shape, plant part, augmentative, made of wood, cohesive mass
7	Artefacts, pejorative, derived
9	Animals, artefacts, loan (other)
11	Narrow things, wavy-shaped things
12	Diminutives
14	Abstract, mass
15	Derived, infinitives

Morphophonological features coded for

Morphophonological feature	Expected class
Mû-	1
Mû-	3
Î-	5
Kî-	7
N, ø	9
Rû-	11
Ka-	12
Û-	14
Kû-	15

Corpus Analysis

- Corpus analysed for rule productivity using Tolerance Principle (Yang, 2016)
- Tolerance Principle demonstrates how to establish the productivity of rules with exceptions.
  - Proposed to account for how and when e.g., language learners make generalizations and when they don't.
  - provides a threshold beyond which the exceptions should not exceed for the rule to be productive.
  - Makes use of 2 integer values N (number of items in the lexicon) and e (number of items not obeying the rule)

  - Let a rule be defined over a set of N items.
  - The rule is productive if the number of exceptions does not exceed a threshold, $\boldsymbol{\theta}_{N.}$

    exceptions (e) $\leq \boldsymbol{\theta}_N (N/\ln(N))$

Corpus Results

Results: a productive semantic feature for class 5

Rule: [+augmentative] ⟶ class 5

Semantic features	N	Class 5	e	θ_N	Productive?
Human	252	2	250	46	No
Augmentative	32	25	7	9	Yes
Round	44	12	32	12	No
Plant part	47	6	41	12	No
Fruit	17	8	9	6	No
Made from wood	12	1	11	5	No
Derived	1081	40	1041	155	No
Cohesive mass	58	5	53	14	No

Corpus results: productive semantic features (overall)

i) [+augmentative] → class 5
ii) [+pejorative] → class 7
iii) [+diminutive] → class 12
iv) [+derived] → class 15
v) [+infinitive] → class 15

Results: productive semantic features

What about [+Human] feature!

Results: a productive morphophonological feature

Rule: [+kî-] ⟶ class 7

Morpho-phonological features	N	Class 7	e	θ_N	Productive?
mû-	397	0	397	66	No
rû-	122	0	122	25	No
kî-	268	262	6	48	Yes
ri-	5	0	5	3	No
kû-	677	0	677	104	No
û-	107	0	107	23	No
î-	188	0	188	36	No
ka	82	0	82	19	No
n-	329	0	329	57	No

Results: productive morphophonological features

i) [+î-/rî-] ⟶ class 5
ii) [+kî-] ⟶ class 7
iii) [+n-] ⟶ class 9
iv) [ø-] ⟶ class 9
v) [+rû-] ⟶ class 11
vi) [+ka-] ⟶ class 12
vii) [+û-] ⟶ class 14
viii) [+kû-] ⟶ class 15

Corpus results: productive morphophonological features

[+Mû-] is not productive!

Are [+human] and [+mû-] really unproductive?

- A possible case of interaction with other features
- A need for recursive application of Tolerance Principle
  - Sub-divide the lexicon into logical subset, reapply the rule

Are [+human] and [+mû-] really unproductive?

- Possible rules based on subsets:
  - [+human, -pejorative] → class 1(semantic)
  - [+human, +pejorative] → class 7 ”
  - [+mû, +human] → class 1 (morphophonology + sematics)
  - [+mû-, -human] → class 3 ”

Results: a productive semantic rule (recursive appl.)

Rule: [+human, -pejorative] ⟶ Class 1

Semantic features	N	Class 1	e	θ_N	Productive?
Human	252	197	52	46	No
[+human, -Pejorative]	235	197	38	43	Yes

- Recursive application yields!

Results: a productive morphophonological rule

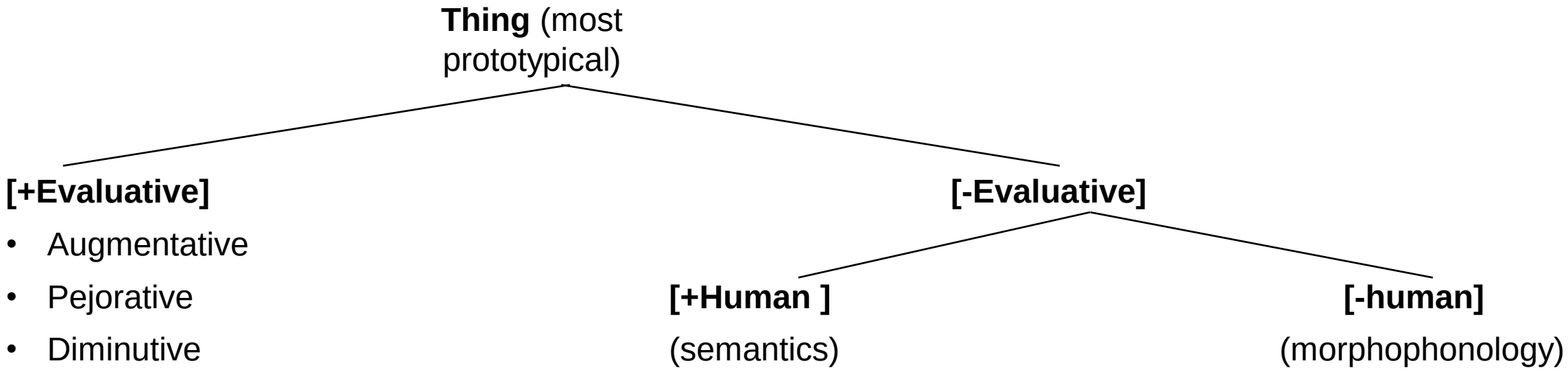
Rule: +mû- [+human] ⟶ Class 1

Hybrid features	N	Class 1	e	θ_N	Productive?
mû-	397	162	235	66	No
[+mû-, +human]	166	162	4	32	Yes

Rule: +[+mû-, -human] ⟶ Class 3

Hybrid features	N	Class 3	e	θ_N	Productive?
mû-	235	235	162	66	No
[+mû-, -human]	235	235	0	43	Yes

A possible schema



Experiments: Observation of human behaviour

- Test to what extent statistical measure of productivity reflects speakers' own usage
  - 3 experiments (two done)
  - 30 native speakers (males & females; age 18—45)

Experiment 1: Semantic features tested

Semantic Feature	Expected Class
Human	1
Extended	3
Augmentative	5
Fruits	5
Artefacts	7
Pejorative	7
loan (other)	9
Narrow	11
Wavy	11
Diminutive	12

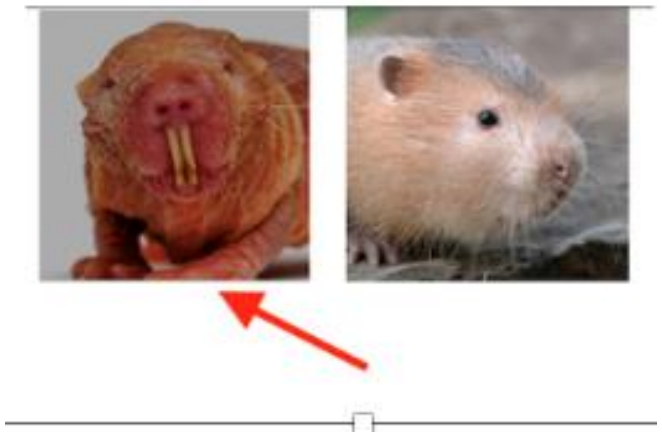
Stimuli

- Participants saw an unfamiliar image followed by a DP(novel noun> num) with an underscore at the beginning of each word
- Required to use their Kĩĩtharaka knowledge to judge if part of the words were missing and type them in full.
- Order in which the stimuli appeared and image-DP pairing randomized across participants
    - Each feature tested 3 times, 30 images, 30 novel nouns

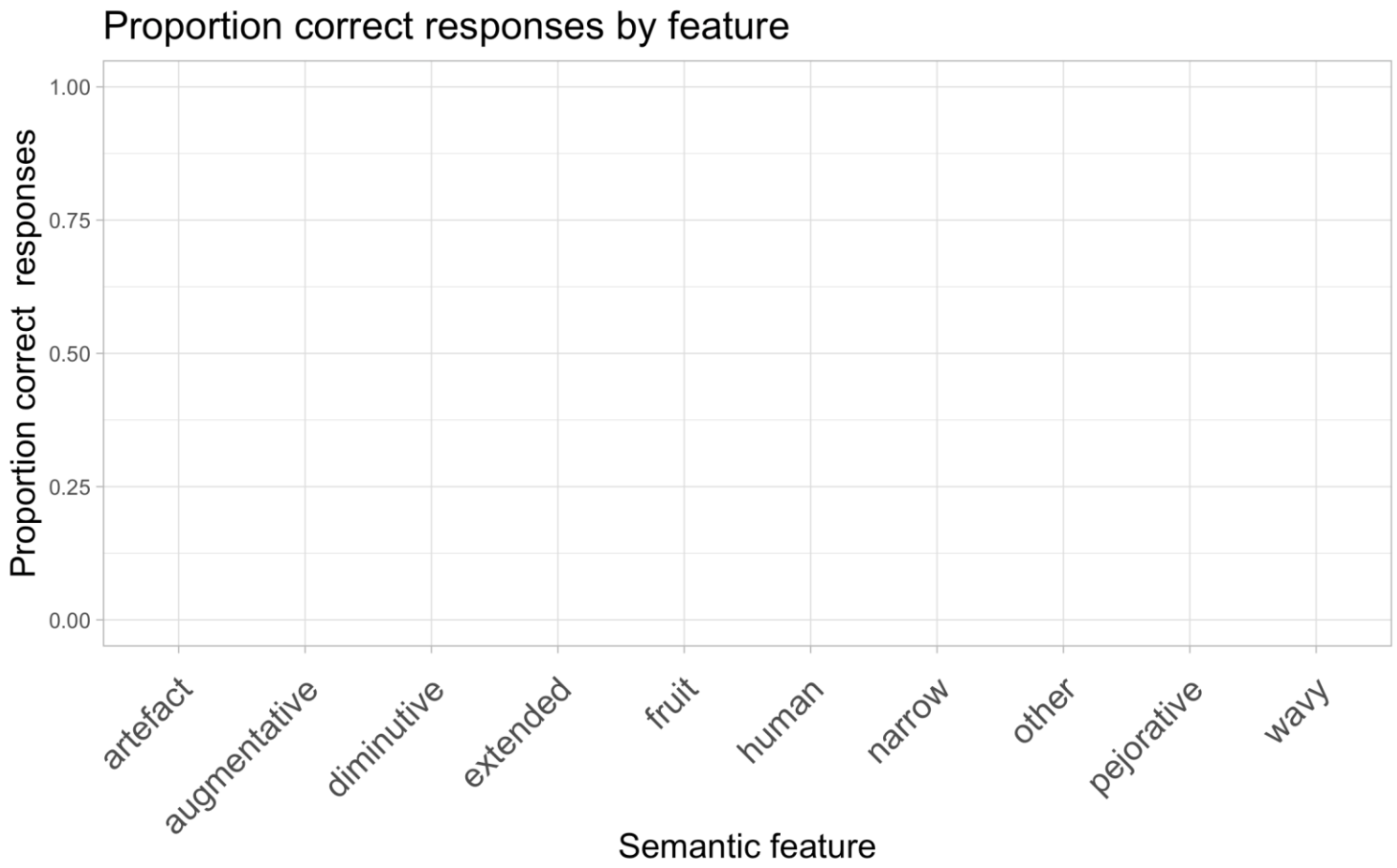
_thindû _mwe

Sample stimuli images

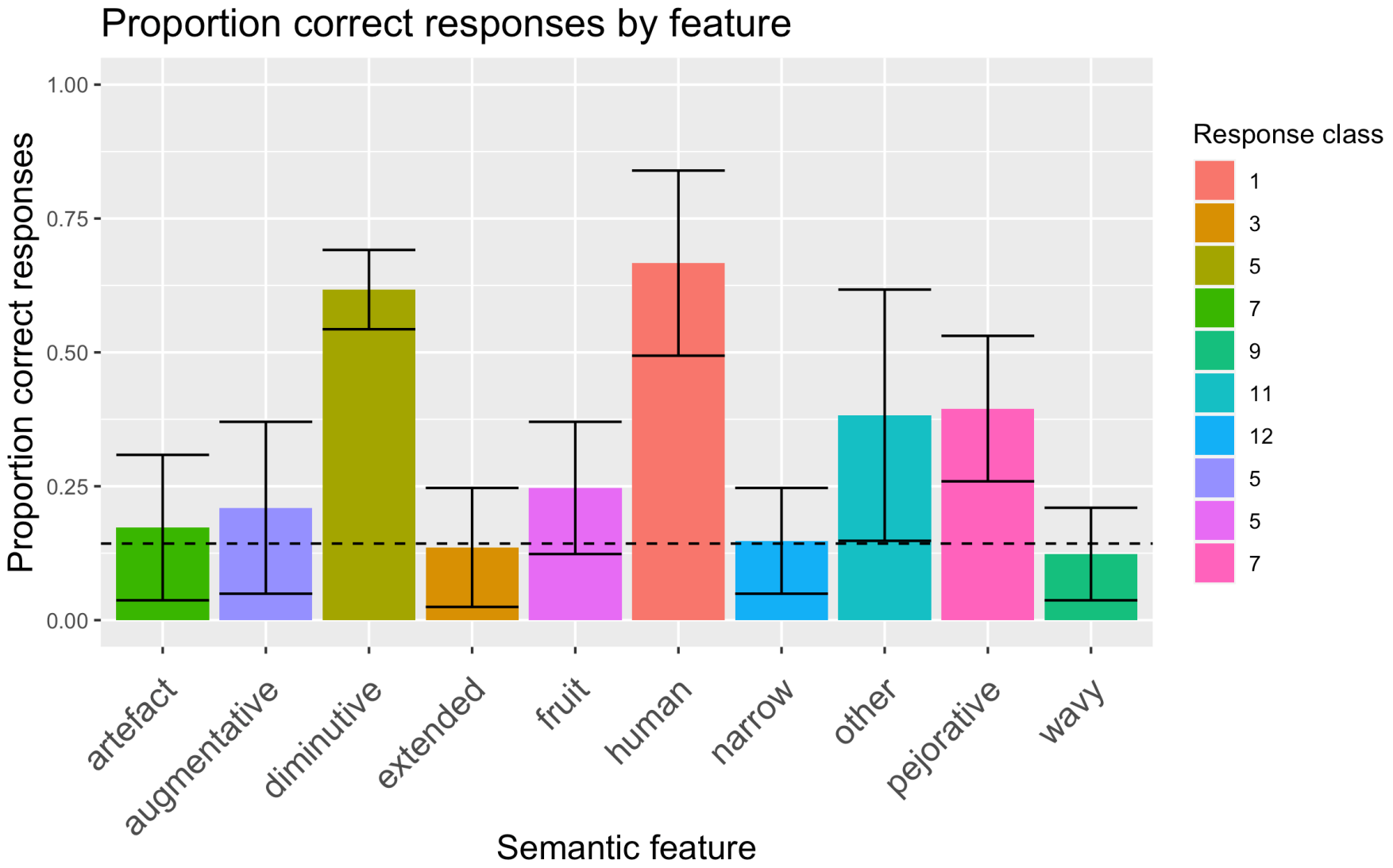
Ngûgî: Kûringa mûtûûrirû wa Gĩĩskoti



Experiment 1 results: productive semantic features



Experiment 1 results: productive semantic features



Does Experiment 1 confirm TP predictions?

- The experimental results largely mirrors the the TP predictions:
  - ¾ features predicted by TP are productive.
  - Augmentative unproductive in the experiment
- Productivity of ‘other’ at chance in the experiment (was unproductive in the corpus)

Experiment 2: Morphophonological features tested

- Prefix-agreement pairing tested in three levels:
  - Singular prefix – singular agreement
  - Plural prefix – Plural agreement
  - Singular prefix – Plural noun + agreement

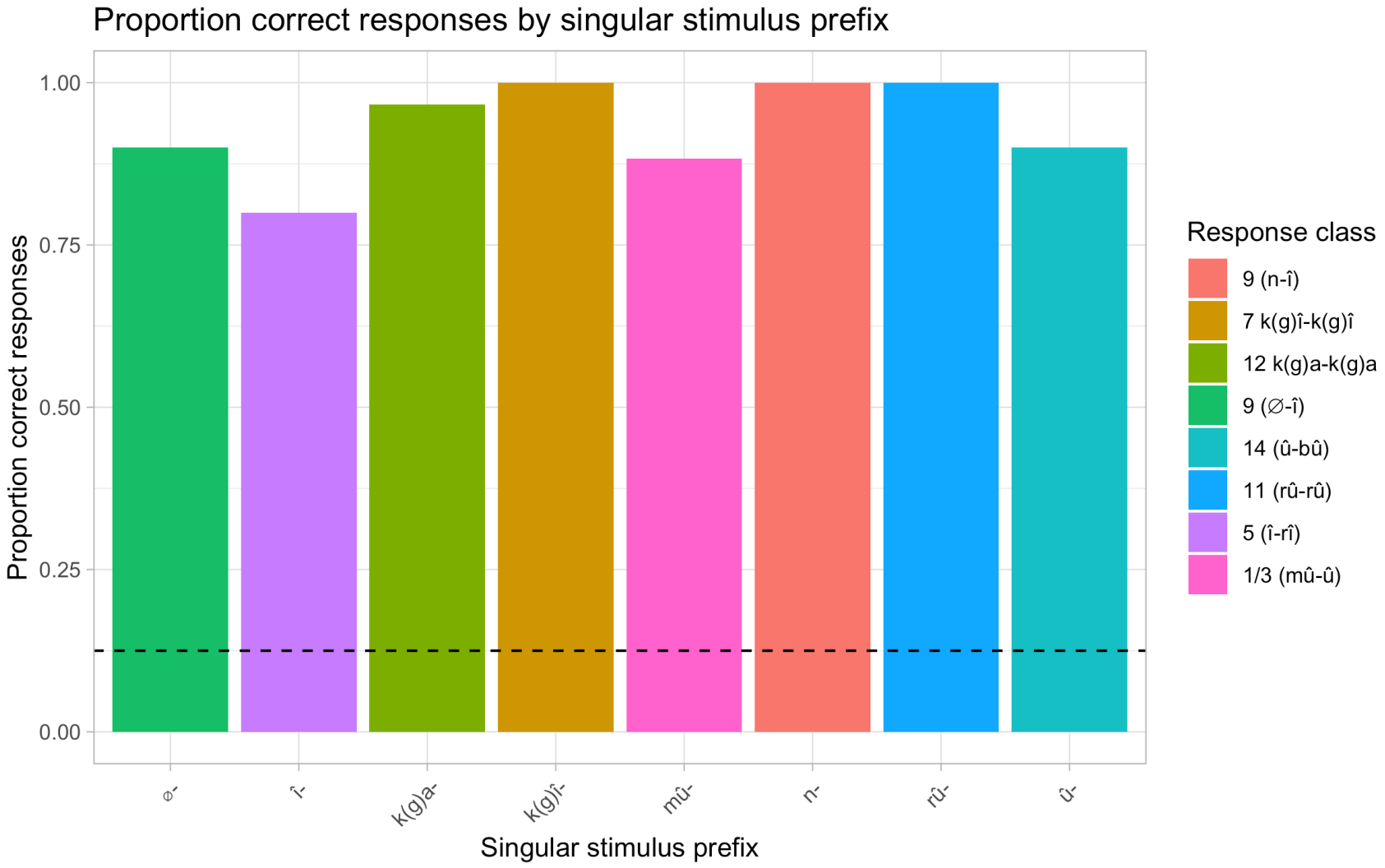
- Participants (30 native speakers, age 18-45)

Stimuli

- Participants saw a (non)prefixed novel noun followed by a short sentence with two gaps, as in *Nkwona* ____ *_mwe* “I have seen _one ____”.
  - Required to fill in the gaps using the noun seen and its agreement

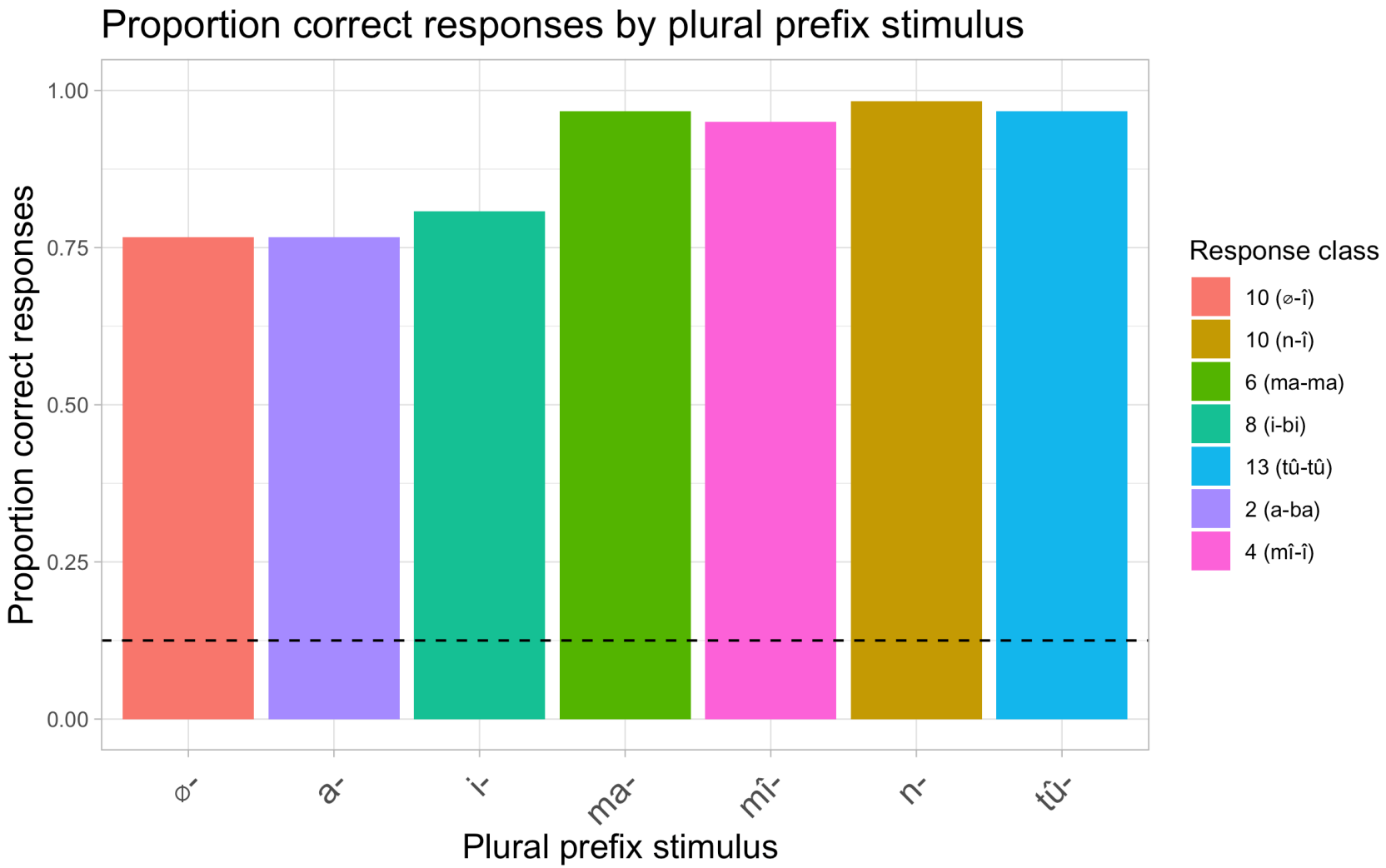
Results: productive morphophonological features

singular stimulus – singular agreement (class)



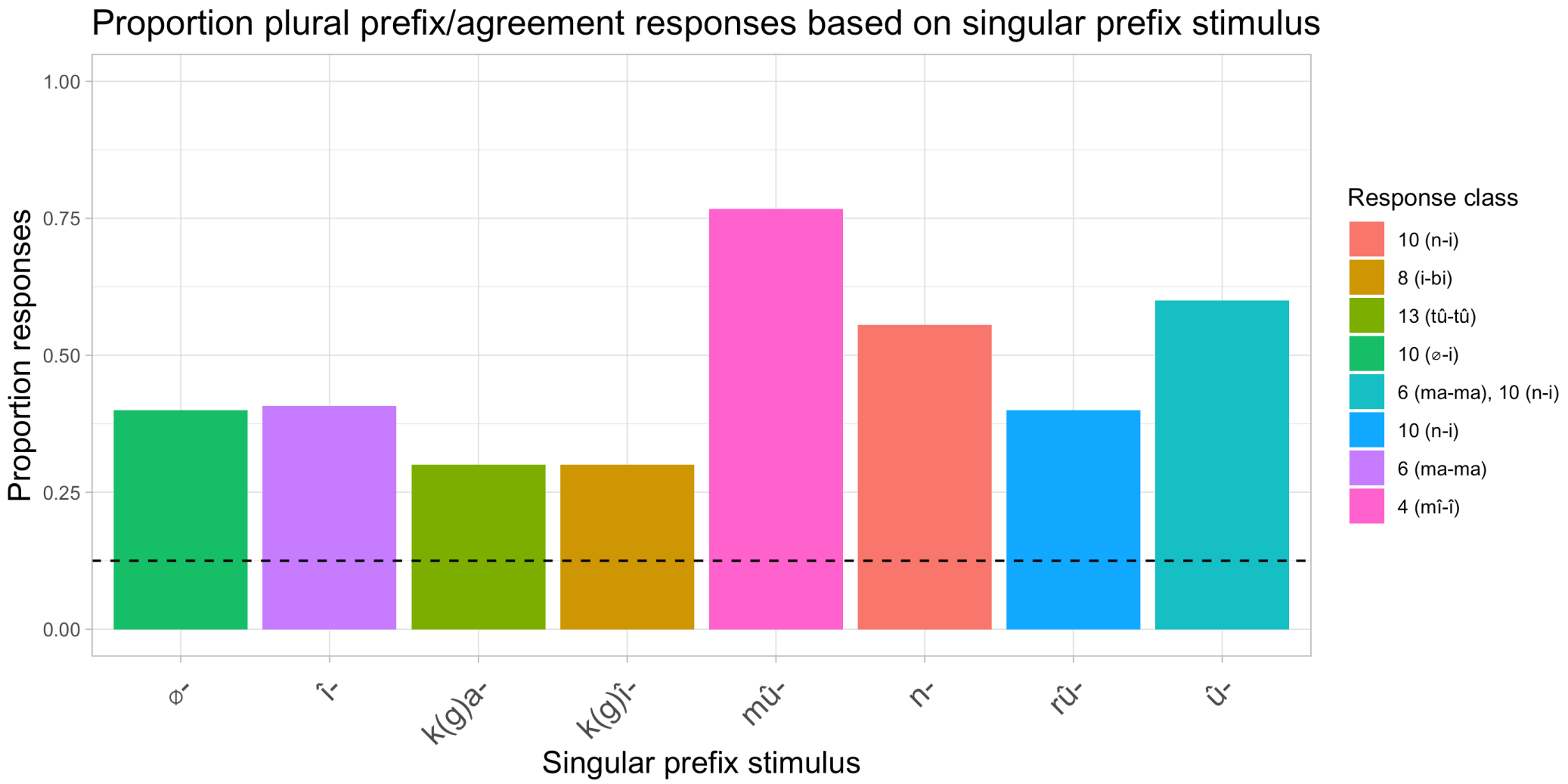
Results: productive morphophonological features

plural stimulus – plural agreement (class)



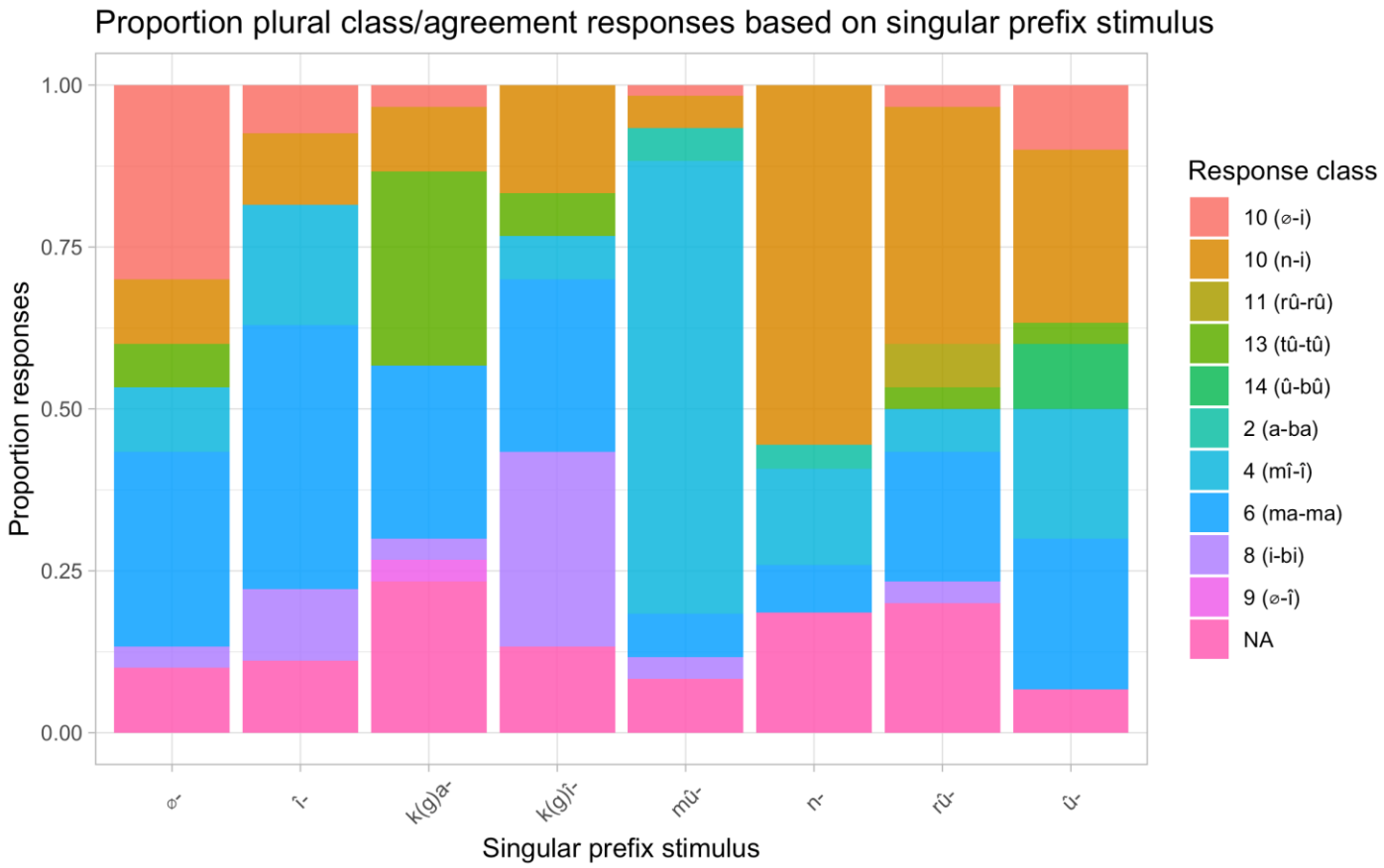
Results: productive morphophonological features

singular stimulus – plural agreement (class)



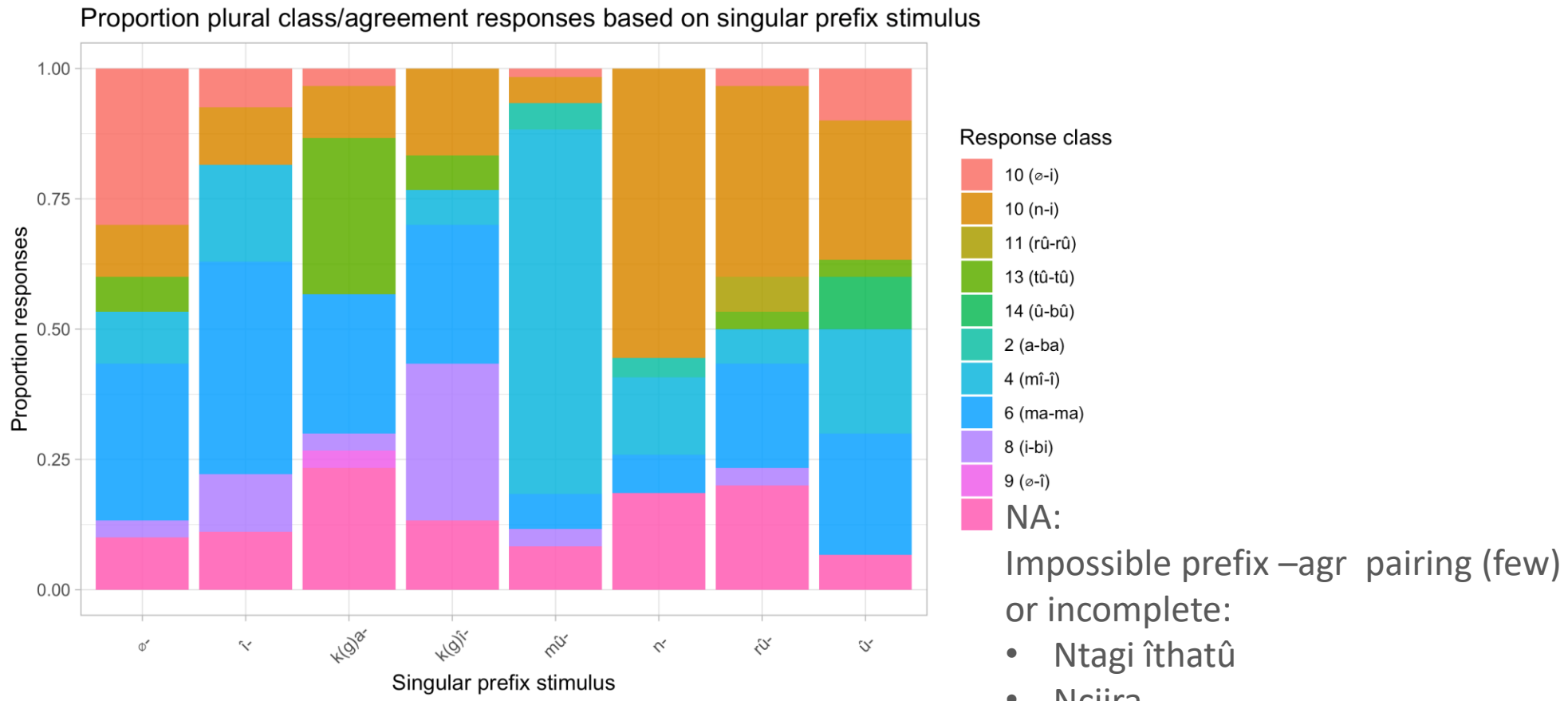
Results: productive morphophonological features

singular stimulus – plural agreement (class)



Results: productive morphophonological features

singular stimulus – plural agreement (class)



Does Experiment 2 confirm TP predictions?

- All prefixes productive in the corpus also productive in the experiment
- The *mû-* prefix is interesting–only productive for class 3
  - Prefix underspecified– but why is it overwhelmingly productive for class 3?
    - Could be a default morphophonological rule that applies in the absence of semantic specification

Implications

- A theory of productivity is instrumental in establishing which features productively determine nominal (gender) agreement
  - Experimental results reflect TP based results to a greater extent.
- Both semantics and formal features are necessary cues to class in Kîîtharaka
  - Semantics- human–nonhuman distinction (class 1/2 vs others); evaluation (class 5, 7, 12)
  - Morphophonology- applies productively for all non-human non-evaluative classes
- Singular-plural class mapping somewhat regular in 1/3, 9 and 14 and less of it in other classes (level 3 exp. 2) – interesting?

Implication: a possible schema

Thing (most prototypical)

[+Evaluative]

Augmentative (î - ma)

Pejorative (k(g)î - bi)

Diminutive (k(g)a-k(g)a

[-Evaluative]

[+Human]

û - ba (1/2)

[-human]

Morphophonological rules

û - î (3/4)

î - ma (5/6)

k(g)î-bî (7/8)

î - i (9/10)

k(g)a - tû (12/13)

û – bû (14)

k(g)û – k(g)û (15)

a - g(k)û (16,17)

Next steps

- Establish relative importance of semantic & morphophonology (experiment 3)
  - Provide stimuli with conflicting semantic and morphophological cues

(following; Karmiloff-Smith, 1981; Gagliardi & Lidz 2014, and others).

The End

Thank you!

References

- Aikhenvald, A. (2003). *Classifiers: A typology of noun categorization devices.* Oxford: Oxford University Press.
- Amidu, A.A. 1997. *Classes in Kiswahili*. Köln: Rüdiger Köppe Verlag.
- Bible Translation & Literacy (EA). (2019). *Bibiria: Kîrîîkanîro kîra gîkûrû na kîrîîkanîro kîra kîyerû.* Nairobi: Bible Translation & Literacy (EA).
- Bleek, W.H.I. (1862-1969). A Comparative Grammar of South African Languages, I: Phonology (1862), II: The Concord, section I: The Noun (1869). London: Trubner. (Republished 1971 in one volume, Farnborough: Gregg.)
- Brooks, P. J., Braine, M. D., Catalano, L., Brody, R. E., & Sudhalter, V. (1993). Acquisition of gender-like noun subclasses in an artificial language: The contribution of phonological markers to learning. Journal of Memory and Language, 32, 76–95.
- Carstens, V., (1991). The morphology and syntax of determiner phrases in Kiswahili. UCLA PhD thesis.
- Corbett, G.G. 1991. *Gender*. Cambridge Mass.: Cambridge University Press.
- Demuth, K. (1992). Acquisition of Sesotho. In Slobin, D. (Ed.), The cross-linguistic study of language acquisition (Vol. 3, pp. 557–638). Hillsdale, NJ: Lawrence Erlbaum.
- Demuth, K. (1994). On the 'underspecification' of functional categories in early grammars. In Lust, B., Suñer, M., Whitman, J. (Eds.), Syntactic theory and first language acquisition: Cross-linguistic perspectives (pp. 119–134). Hillsdale, NJ: Lawrence Erlbaum.
- Demuth, K. (2000). Bantu noun class system: Loan words and acquisition. (G. Senft, Ed.) *Classification Systems*, 270-292.
- Demuth, K., & Ellis, D. (2008). Revisiting the acquisition of Sesotho noun class prefixes. Mahwah, NJ: Lawrence Erlbaum.
- Denny, J.P. & Creider, C.A. 1986. The semantics of noun classes in Proto-Bantu. In *Noun classes and categorization*. C. Craig, Ed. Amsterdam/ Philadelphia: John Benjamins Publishing Company.
- Desrochers, A., Paivio, A., & Desrochers, S. (1989). L'effet de la fréquence d'usage des noms inanimés et de la valeur prédictive de leur terminaison sur l'identification du genre grammatical. *Revue Canadienne de Psychologie,* **43,** 62–73.
- Dixon, R. (1986). Noun classes and noun classification in typological perspective. In C. G. Craig (Ed.), Noun classes and categorization (pp. 105–112). Philadelphia, PA: John Benjamins.
- Eberhard, David M, F. Simons and Charles, D. (2019). *Kîîtharaka.* (Fennigh, Ed.) Retrieved 09 28, 2019, from Ethnologue: Languages of the World: http://www.ethnologue.com
- Frigo, L., & McDonald, J. (1998). Properties of phonological markers that affect the acquisition of gender-like subclasses. Journal of Memory and Language, 39, 218–245.
- Gagliardi, A. and Lidz, J. (2014). Statistical insensitivity in the acquisition of Tsez noun classes. Language, 90(1):58–89..
- Givón, T. (1971b). Historical syntax and synchronic morphology: an archaeologist's field trip. *CLS*. 7:394–415.
- Guthrie, Malcolm (1967–71) *Comparative Bantu: an introduction to the comparative linguistics and prehistory of the Bantu languages*. 4 vols. Farnborough: Gregg Press.
- Hockett, C. F. (1958) A Course in Modern Linguistics. New York: Macmillan. 1966. What Algonquian is really like. *International Journal of American Linguistics* 32:59-73
- Msaka, P.(2019) Nominal classification in bantu revisited: The perspective from Chichewa. PhD dissertation: *Stellenbosch University. Available at:* http://hdl.handle.net/10019.1/107220

Derivation of the Tolerance Principle

TP assumes that the distribution of lexical items follow Zipf's law (Zipf, 1949):

- The rank (r) of a word in a naturally occurring corpora is inversely proportional to its frequency. Using this assumption, we can calculate the probability of a random ranked word, thus:

$$p_i = f_i / \sum_{N=1}^{N} fk$$

$$= (\frac{C}{r_i}) / \sum_{k=1}^{N} \frac{C}{r_k}$$

$$= \frac{1}{iH_N} \; where \; H_N = \sum_{k=1}^{N} \frac{1}{k}$$

(Yang, 2018:2)

For a non-productive rule (T(N, N)), rank = r = r units of time, hence:

Sum probability = $\sum_{r=1}^{N} r \frac{1}{,HN} = \frac{N}{H_N}$