Determinants of noun classification in Kĩĩtharaka: A synchronic approach

Patrick Kanampiu[1], Alexander Martin[2], Jennifer Culbertson[1]

[1]University of Edinburgh, [2]University of Paris

9th Conference on Bantu Linguistics

7th -10th June,2022

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** camionn**ette**
Det.F van.F

What determines noun class in a language?

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

Kĩĩtharaka

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

(2) **kî**banga **kî**ra ← morpho-phonology (kî-kî/*kî/i*)
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 (*-ette*)
Det.F van.F

What determines noun class in a language?

- Discovering exact set of features that determine noun classification in a given language is an intricate affair!
- Typically, many exceptions…

Kĩĩtharaka

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

(7) **kî**roria **kĩî**ra ← ?
7.prophet 7.dem.dist

(8) **î**the **mû**kûrû ← ?
5.father 1.elder

French

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

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

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

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

Theoretical and empirical questions…

- How do we know what semantic or morpho-phonological 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)

Required:

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

- Have these questions been answered by traditional approaches to Bantu noun classes?

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	**a**thi
	9.devil	1.went

- Centred on semantics – problematic (whether meaning is on the nominal stem or prefixes)

Traditional approaches to Bantu noun classes

- No theory for saying what makes something productive!
- No evidence that speakers find these things productive!
- A different approach:
  - quantitative corpus (e.g., Ngcoco, 2010 (Zulu); Msaka, 2019 (Chichewa).
  - psycholinguistic experiments with speakers (e.g., Jarro et al, 2021 (Kinyarwanda).
  - Ideally both corpus and psycholinguistic experiments (the current study).

A different approach to Kĩĩtharaka

- The study aims to establish productive rules that determine noun class membership in Kĩĩtharaka:
  - Corpus analyses using Tolerance Principle (Yang, 2016) (plus native speaker intuitions) provides predictive rules for class assignment.
  - Run psycholinguistic tests – will show 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 – Currently (2328 nouns)
  - ~1000 collected from Kĩĩtharaka bible and,
  - ~1300 translated from SIL comparative African Wordlist (Snider and Roberts, 2006).
- Coded for singular/plural agreement, semantics and morphophonological features (prefixes).

Corpus Analysis

Corpus analysed for rule productivity using Tolerance Principle (TP) (Yang, 2016)

- Tolerance Principle demonstrates how to establish the productivity of rules with exceptions.
  - 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, $\boldsymbol{\theta}_{N.}$

$$\text{exceptions} \leq \boldsymbol{\theta}_N = N/\ln(N) \qquad \text{(Yang, 2016)}$$

Semantic features coded for

TARGET CLASS	SEMANTIC FEATURE(S)
1/2	(super)human
3/4	Extended shape, spread shape, protruding shape, trees & plants, dispersive mass)
5/6	Fruits, round shape, plant part, augmentative, made of wood, cohesive mass
7/8	Artefacts, pejorative, derived
9/10	Animals, artefacts
11/10	Narrow things, wavy-shaped things
12/13	Diminutives
14	Abstract, mass
15	Derived, infinitives

Morphophonological features coded for

TARGET CLASS	MORPHOPHONOLOGICAL FEATURE(S)
1/2	M(u)û-
3/4	Mû-
5/6	Î-
7/8	Kî-
9/10	Mb-/Mp-/Nd-/ Nt-/Nth-
11/10	Rû
12/13	Ka-
14	U-
15	Kû-

Results: a productive semantic feature

Semantic features	N	Class 5	e	θ_N	Productive?
Human	254	2	252	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

Results: unproductive semantic feature

Semantic features	N	Class 3	e	θ_N	Productive?
Animal	155	15	140	31	No
Protruding	16	3	13	6	No
Tree/plant	55	25	30	14	No
Artefact	136	32	104	28	No
Narrow	115	32	83	24	No
Derived	1081	33	1048	155	No
Cohesive	58	10	48	14	No
Dispersive	64	13	51	15	No
Extended	161	82	79	32	No
Spread	49	1	48	13	No

Results: a productive morpho-phonological feature

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: unproductive morpho-phonological feature

Morpho-phonological features	N	Class 1	e	θ_N	Productive?
mû-	397	162	235	66	No
rû-	122	0	122	25	No
kî-	268	0	268	48	No
ri-	5	0	5	3	No
kû-	677	0	677	104	No
û-	107	0	107	23	No
î-	188	1	187	36	No
ka-	82	0	82	19	No
n-	329	8	321	57	No

Results: Productive semantic rules

i) If [+human] ⟶ class 1/2

ii) If [+augmentative] ⟶ class 5/6

iii) If [+human] + [pejorative] ⟶ class 7/8

iv) If [+ diminutive] ⟶ class 12/13

v) If [+derived] ⟶ class 15

vi) If [+infinitive] ⟶ class 15

Results: Unproductive semantic rules

i) If [+animal] ⟶ class 3, 5, 7, 9, 11

ii) If [+plant part] ⟶ class 5, 11

iii) If [+wooden] ⟶ class 5

iv) If [+narrow] ⟶ class 11

v) If [+wavy] ⟶ class 11

vi) If [+artefacts] ⟶ class 7

vii) If [+protruding] ⟶ class 3

viii) If [+dispersive mass] ⟶ class 3

ix) If [+cohesive mass] ⟶ class 5

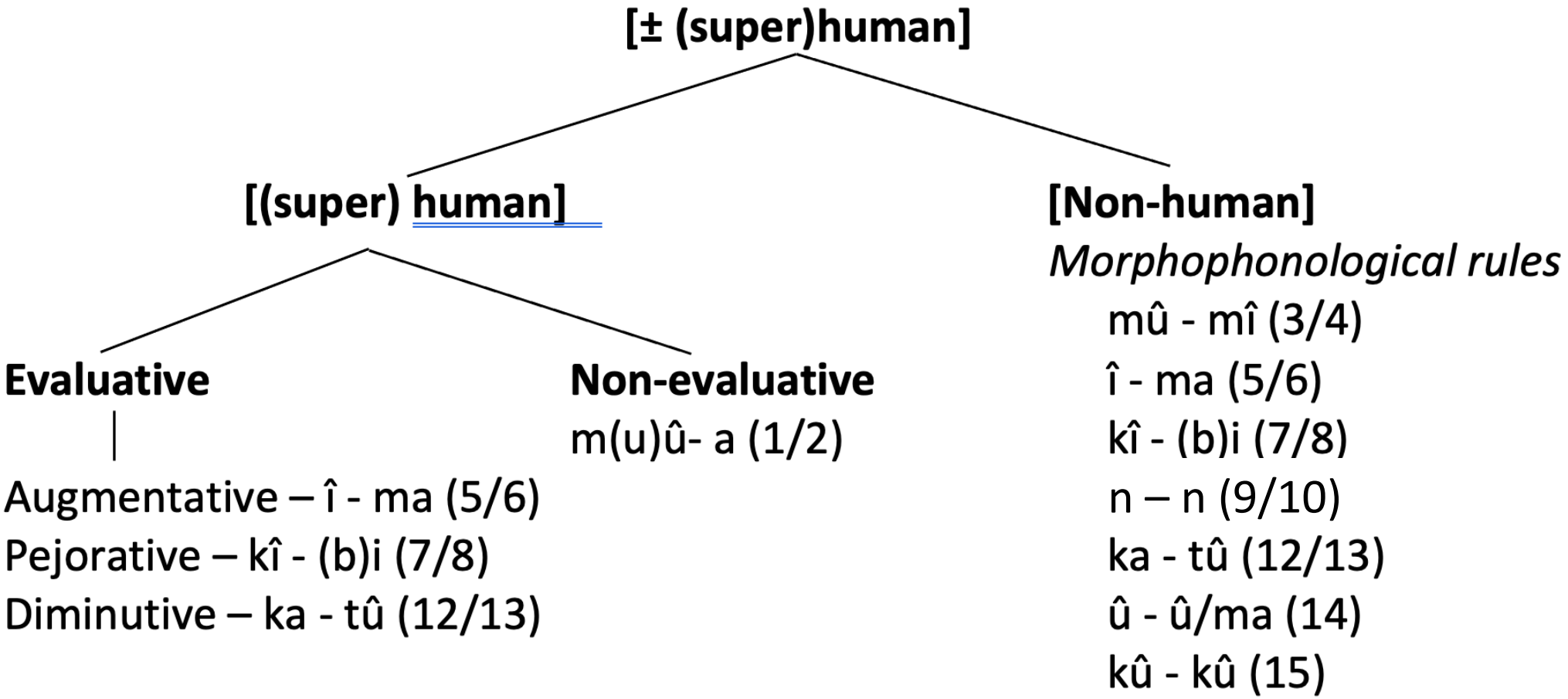
Results: Productive morpho-phonological rules

i) If [[+human] + [mû-]] ⟶ class 1/2
ii) If [[-human]+ [mû-]] ⟶ class 3/4
iii) If [+î-/rî-] ⟶ class 5/6
iv) If [+kî-] ⟶ class 7/8
v) If [+n-] ⟶ class 12/13
vi) If [+rû-] ⟶ class 11
vii) If [+û-] ⟶ class 14
viii) If [+kû-] ⟶ class 15

Implications

- Most semantic features predicted to be unproductive under Tolerance Principle
- Most morphological features predicted to be productive
  - class 1/2, 3/4 have exceptions
- [±(super)human] a prime semantic feature in classification of Kĩĩtharaka nouns

Implications: a possible schema



Implications: a possible schema

		Thing (most prototypical)			
Evaluative				Non-evaluative	
			(Super)Human		Non-(super)human Morphophonological rules
					û-î (3/4)
					Î -ma (5/6)
					K(g)î-k(g)î (7/8)

Implications: comparison with Chichewa

Two super-agreement noun classes (Msaka, 2019)

a) human/animate/agentive ~ animate

b) non-human/inanimate/non-agentive ~ inanimate

The semantically-determined super-ANC in Chichewa

Animate nouns		**Inanimate nouns**	
m(u)-a-ANC	(≈NC1)	*li*-ANC	(≈NC5)
		i-ANC	(≈NC4 and 9)
		u-ANC	(≈NC3 and 14)
		a-ANC	(≈NC2 and 6)
		chi-ANC	(≈NC7)

Corpus Limitations

- Limited variety of written Kîîtharaka text- relied on Kîîtharaka bible (BTL, 2019)
  - Bible lacked diversity on some semantic domains (e.g., names of local trees, local animals, body parts etc)
  - Translated SIL comparative African word list (Snider & Roberts, 2006)
- Tolerance Principle is a mere prediction of human behaviour
  - Need for confirmation through experimenting with humans.

Next steps: Psycholinguistic experiments

- **Goals:**
  - Test TP predictions
    - What semantic and morpho-phonological rules are speakers sensitive to when classifying novel nouns?
  - What is the relative importance of semantic and morpho-phonological cues?
    - Results of similar experiments outside Bantu: children learning noun classes prefer phonology cues even when statistically less reliable than semantic cues (as in Karmiloff-Smith, 1981; Gagliardi & Lidz, 2014 a.o.)

Conclusion

- Noun classes are hard to understand…
  - If approached from a diachronically centric point of view
  - Without a theory of productivity, and without empirical evidence from speakers
- A new approach that makes use of data from synchronic corpus and observation of human behaviour can help us better understand Kîîtharaka noun classes

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

Results: a productive semantic feature (class-feature mapping)

Class	Human	N	e	θ_N	Productive?
1	198	198	0	37	Yes
3	4	223	219	41	No
5	2	177	175	34	No
7	21	262	241	47	No
9	22	428	406	71	No
11	0	120	120	25	No
12	6	81	75	18	No
14	0	105	105	23	No
15	0	684	684	105	No

Results: a productive morpho-phonological feature (class-feature mapping)

Class	Prefix	N	e	Theta_N	Productive?
1	162	198	36	37	Yes
3	223	223	0	41	Yes
5	172	177	5	34	Yes
7	261	262	1	47	Yes
9	337	428	91	71	No
11	120	120	0	25	Yes
12	81	81	0	18	Yes
14	105	105	0	23	Yes
15	675	684	9	105	Yes