A Corpus Study of
Swahili's Dual-Complementizer System

Aron Finholt & John Gluckman

University of Kansas

finhola@ku.edu
johnglu@ku.edu

April 8, 2021

Distribution of *kwamba* and *kuwa*

This project investigates the distribution of the Swahili complementizers *kwamba* and *kuwa*, which are both used to introduce a finite indicative clause under clause embedding predicates like *-ambia*, 'tell' (1).

(1)   *Hamisi a-li-ni-ambia*   **kwamba/kuwa** *a-na-penda*   *kusoma*
      Hamisi 1SM-PAST-1SG.OM-tell COMP/COMP   1SM-PRES-like read.INF
      'Hamisi told me that he likes to read.'   (Mpiranya, 2015:220)

Kuwa and *kwamba* are reported to be interchangeable, with no interpretive or distributional distinctions.

Distribution of *kwamba* and *kuwa*

Although such a distributional description of *kwamba/kuwa* is generally accepted, the fact that these two complementizers appear to exist in free variation is prima facie surprising.

- Each have distinct lexical origins as infinitival verb forms; *kwamba* being derived from 'to tell', and *kuwa* from 'to be'.

- Interpretative differences have been reported in similar dual-complementizer systems (e.g. Greek, Italian Dialects) (Ledgeway, 2000; Angelopoulos, 2019)

- Many Bantu languages have multiple complementizers—including 'say'-complementizers—serving various "evidential"-like functions (Botne, 1997; Güldemann, 2008; Diercks, 2013)

Speaker intuitions

A meaningful distinction between *kwamba* and *kuwa* is moreover motivated by native speaker intuitions.

- *kwamba* feels "weaker" or more "subjective."
- *kuwa* feels "stronger" or more "factual."

The difference is, at best, extremely subtle, and varies greatly depending on the person and the context (and possibly the dialect of Swahili).

- We examine the question of complementizer choice through regression-based analysis of (Tanzanian) Swahili Corpus data.

- Ultimately, we find that complementizer choice in Swahili is (at least) partially predictable based on a subset of factors that have been shown to influence complementizer choice cross-linguistically.

- Based on these results, and taken together with the native speaker judgments, we propose a system in which complementizer choice in Swahili encodes *relative belief*.

Project Overview
oooo

Methods and Results
●ooooooo

Discussion
ooooooo

References

Methodology

Corpus:	Helsinki Corpus of Swahili 2.0
	→ approx 25 million words
	→ fully morphologically tagged
Token Type:	Embedding Predicate+[$_{CP}$ *kwamba/kuwa*. . .]
Total Token Count:	26,064

The factors we investigated were chosen based on what was feasible to look at in a corpus.

Factor 1: Predicate class

Predicate class (or selection) is known to affect complementizer choice cross-linguistically (Kiparsky and Kiparsky, 1971; Hooper and Thompson, 1973; Noonan, 2007; Roussou, 2010), many others

- We initially divided up the predicates based on the classification in Hooper and Thompson (1973).
- Eventually, we collapsed these into just two categories
    - **Attitude predicates** (-*fikiri* 'think'): those predicates which entail the existence of a belief-holder.
    - **Reportative predicates** (-*sema* 'say'): those predicates which do not entail the existence of a belief-holder.

Factor 2: Person of subject

The person of the main-clause subject has also been shown to affect
complementizer choice cross-linguistically (Kiparsky and Kiparsky, 1971; Givon and
Kimenyi, 1974; Massamba, 1986)

- In Kinyarwanda, the complementizer *kongo* is reported to not be possible
  under factives with 1st/2nd person subjects (Givon and Kimenyi, 1974).

(2)   a.   *yiibagiwe   kongo amazi yari mare-mare*
             3SG.forgot COMP water was deep

             'He forgot that the water was deep (and I doubt it).'

       b.   * *niibagiwe   kongo amazi yari mare-mare*
             1SG.forgot COMP water was deep

             [Intended: 'I forgot that the water was deep (and I doubt it).']
             (Givon and Kimenyi, 1974, 110)

Factor 3: Mood

The mood of the embedded clauses has also been shown to be a factor in complementizer selection (Ledgeway, 2000; Roussou, 2010), many others.

- Subjunctive mood in the embedded clause is often correlated with a particular complementizer as in, e.g., Greek.

Project Overview
oooo

Methods and Results
ooooo●ooo

Discussion
ooooooo

References

Overview of results

Factors investigated

1. Predicate class
2. Person of subject
3. Mood

- All three factors are significant predictors of complementizer choice

- *However*, the person of the main clause subject is by far the most significant factor affecting complementizer selection.
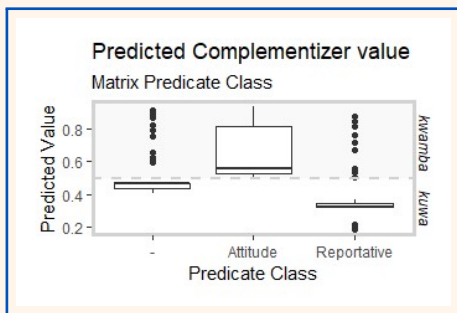
Results: Matrix Subject Person

Predicted Complementizer Value
Matrix Subject Person
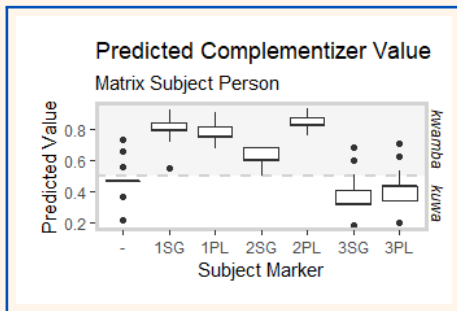
Key Findings

- Matrix Subject Person found to be the strongest individual predictor in the model.

- Only *first*-person subjects were shown to be a statistically significant predictor; correlate with *kwamba*.
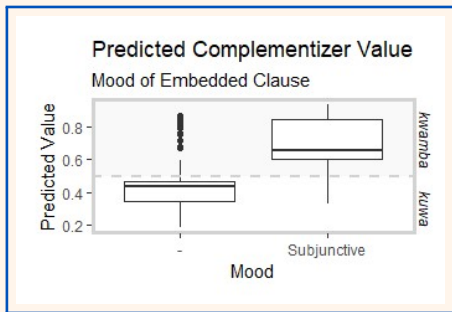
Results: Matrix Predicate Class

Key Findings

- Matrix Predicate Class found to be second strongest predictor in the model.

- ATTITUDE predicates (e.g. *-fikiri*, 'think') shown to correlate with *kwamba*, while REPORTATIVE predicates (e.g. *-sema*, 'say') instead correlate with *kuwa*.

Project Overview
0000

Methods and Results
0000000●

Discussion
0000000

References

Results: Mood of Embedded Clause

Key Findings

- Mood of Embedded Clause found to be weakest predictor in the model.

- Presence of the subjunctive in the embedded clause (e.g. FV ,-*e*) shown to correlate with *kwamba.*

Discussion

We suggest that *kwamba* and *kuwa* generally distinguish between "specific" vs. "general" belief.

- *Kwamba* indicates that the embedded proposition is believed "true for someone"—typically the subject.

- *Kuwa* indicates that the embedded proposition is believed "true for everyone (who is relevant)."

Project Overview
0000

Methods and Results
00000000

Discussion
0●00000

References

Matrix Subject Person

The model identifies *first*-person subjects as being significantly predictive of complementizer choice.

- Specifically, *first*-person is shown to predict the use of *kwamba*.

We interpret this correlation as evidence that with *kwamba*, the embedded proposition is evaluated relative to the local attitude holder.

- The speaker is more aware of their own thoughts than of others' →
  kwamba > *kuwa* in the presence of a 1st person subject.

Matrix Predicate Class

The model distinguishes ATTITUDE and REPORTATIVE predicates.

- *Kwamba* correlates with ATTITUDE predicates, while *kuwa* correlates with REPORTATIVE predicates.

We similarly interpret this dichotomy as evidence that *kwamba* and *kuwa* differ in the individual (or set thereof) that evaluates the embedded proposition.

- We suggest that *kwamba* relativizes the truth of P to the local subject; *kuwa* presents more general knowledge.

- This may also account for the fact that the subjunctive patterns with *kwamba*.

Project Overview
oooo
Methods and Results
oooooooo
Discussion
oooo●ooo
References

The corpus study matches native speaker judgments from at least one Tanzanian.

(3) *Mimi ni-na-jua* **kwamba/kuwa** *Tanzania i-ta-shinda*
1SG 1SG-PRES-know COMP/COMP 9.Tanzania 9SM-FUT-win

'I know that Tanzania will win.'

- With *kwamba*, there is a sense that this is a belief only I can (reasonably) subscribe to; Tanzania is up 1-0 at the half.

- *Kuwa* instead signals that the embedded proposition is (again, reasonably) obvious to all; Tanzania is leading 5-0 with 1 minute left in the game.

Summary of findings

Complementizer	Predictor
kwamba	*first*-person
	Attitude Predicates
	Subjunctive
kuwa	**Reportative Predicates**

Project Overview
0000

Methods and Results
00000000

Discussion
0000●0

References

Conclusion

- Taken with speaker judgement data, the results of our corpus analysis suggest that the free-variation analysis of *kwamba/kuwa* may not be sufficient.

- Based on the factors shown to predict *kwamba*, we suggest that *kwamba* conveys that the embedded proposition is interpreted relative to the local subject (or attitude holder).

- The evidence puts Swahili in line with other Bantu languages which make similar epistemic distinctions in the C domain. It is unclear to us whether the Swahili distinctions were introduced through borrowing, or arose naturally.

Thanks!

Angelopoulos, N. (2019). *Complementizers and Prepositions as Probes: Insights from Greek*. PhD thesis, UCLA.

Botne, R. (1997). Evidentiality and epistemic modality in Lega. *Studies in Language*, 21(3):509–532.

Diercks, M. (2013). Indirect agree in Lubukusu complementizer agreement. *Natural Language & Linguistic Theory*, 31(2):357–407.

Givon, T. and Kimenyi, A. (1974). Truth, belief and doubt in Kinyarwanda. *The Papers from the Fifth Annual Conference on African Linguistics*, 5:95–114.

Güldemann, T. (2008). *Quotative Indexes in African Languages: A Synchronic and Diachronic Survey*, volume 34 of *Empirical Approaches to Language Typology*. Mouton de Gruyter.

Hooper, J. B. and Thompson, S. A. (1973). On the Applicability of Root Transformations. *Linguistic Inquiry*, 4(4):465–497.

Kiparsky, P. and Kiparsky, C. (1971). Fact. In Steinberg, D. and Jakobovits, L., editors, *Semantics: an interdisciplinary reader in philosophy, linguistics, and psychology*. Cambridge.

Ledgeway, A. (2000). *A Comparative Syntax of the Dialects of Southern Italy: A Minimalist Approach*. Blackwell.

Massamba, D. P. B. (1986). Reported speech in swahili. In Coulmas, F., editor, *Direct and Reported Speech*, pages 99–120. Mouton de Gruyter.

Noonan, M. (2007). Complementation. In Shopen, T., editor, *Language Typology and Syntactic Description*, volume II, pages 52–150. Cambridge University Press.

Roussou, A. (2010). Selecting complementizers. *Lingua*, 120:582–603.