Understanding the loss of Proto-Tupí-Guaraní object marking

A case study in Mbyá

Guillaume Thomas

University of Toronto

5th Symposium on Amazonian Languages
University of New Mexico

March 30, 2024

Overview

Proto-Tupí-Guaraní transitive verbs had a prefix that Jensen (1987, 1990, 1998) and others identified as an object marker.

Object marking has disappeared from some but not all Tupi-Guaraní languages.

Jensen (1998) notes that this disappearance “does not appear to have any connection with other cross-referencing changes”

Overview

The goal of this study is to elucidate the factors that are responsible for the loss of object marking in Tupi Guarani languages.

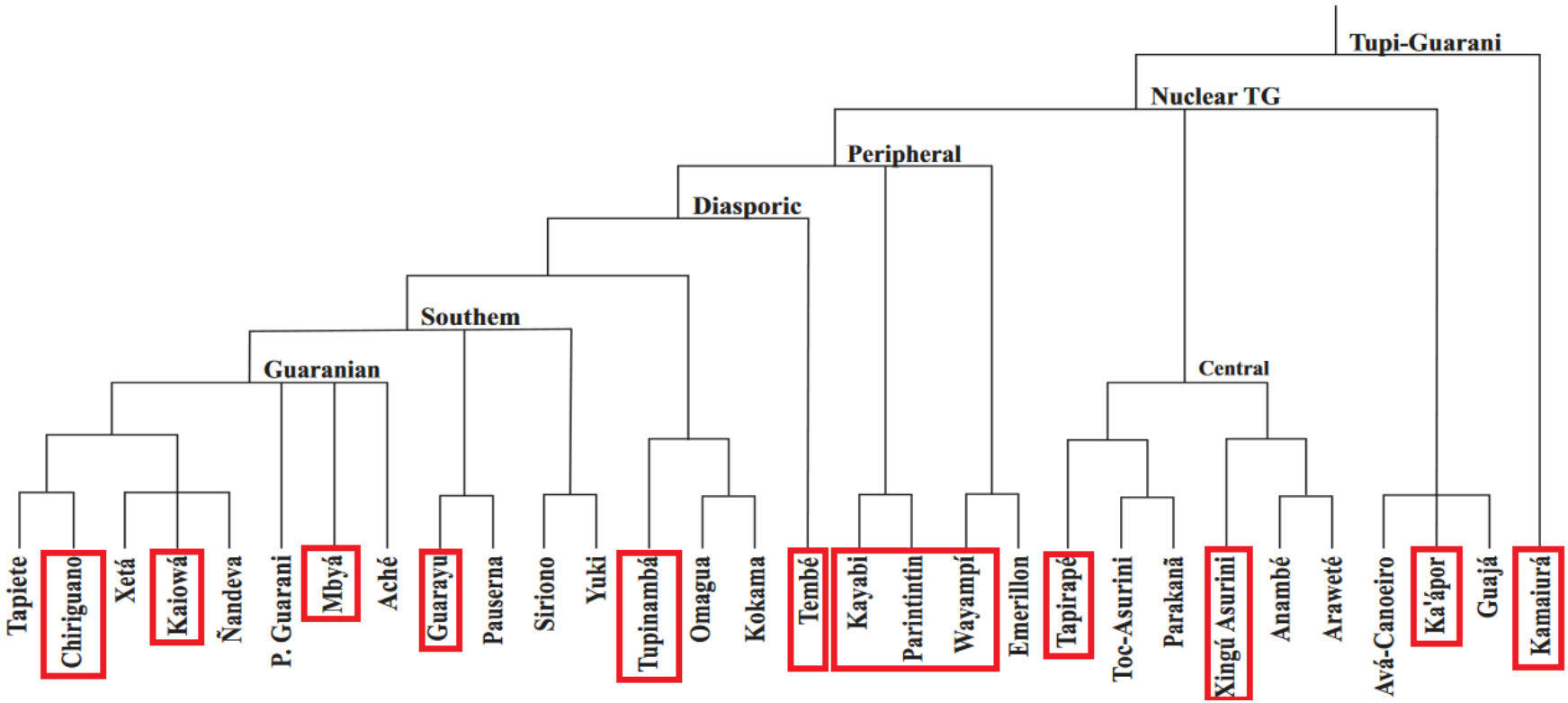
Why is this change unrelated to other cross-referencing changes?

The strategy is to carry out a case study of object marking in Mbyá Guarani:

- In Mbyá, object marking is present with a subset of transitive verbs.
- This makes it possible to study which features of transitive verbs are associated with the loss of object marking.

Loss of PTG object marking from PTG

Tupí-Guaraní languages



Michael et al. 2015

PTG object marking

PTG active-inactive marking (Jensen 1998):

(1) a. *a-manõ — *active*
‘I die’

b. *o-manõ
‘They/she/he/it dies’

(2) a. *čé katú — *inactive class 1*
‘I am good’

b. *i-katú
‘They/she/he/it are/is good’

(3) a. *čé r-orýƀ — *inactive class 2*
‘I am happy’

b. *c-orýƀ
‘They/she/he/it are/is happy’

PTG object marking

PTG hierarchical indexing and object marking (J. 1987, 90, 98):

(4) a. *a-**i**-potár — *transitive class 1*
(‘I want them/her/him/it’)

b. *čé potár
(‘They/she/he/you want(s) me’)

(5) a. *a-**c**-epják — *transitive class 2*
(‘I see them/her/him/it’)

b. čé r-epják
(‘They/she/he/you see(s) me’)

PTG object marking

**jo(c)* OM allomorph with monosyllabic stems (J. 1987, 90, 98):

(6) *a-**jo**-’ók *transitive class 1*
(‘I dig them/her/him/it up’)

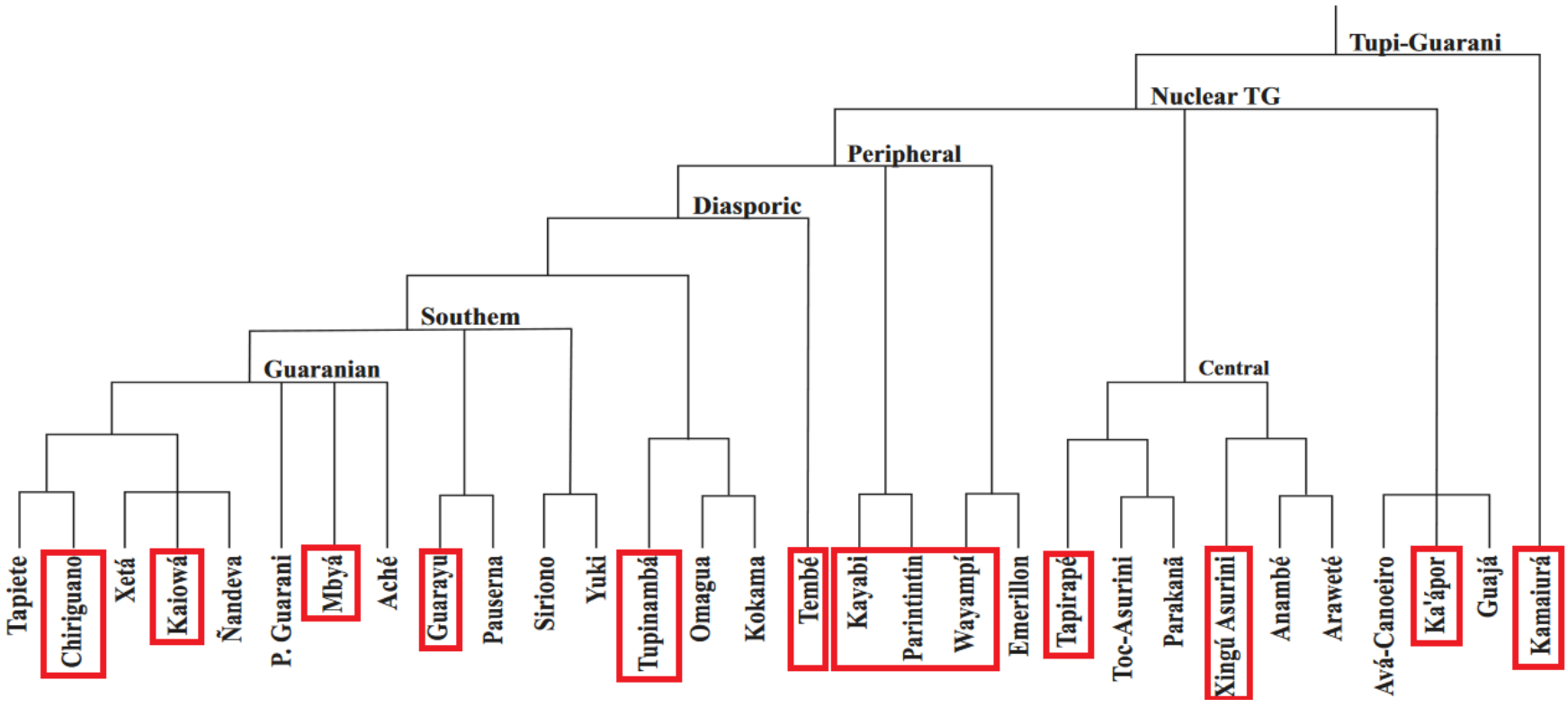
(7) *a-**joc**-éj *transitive class 2*
(‘I wash them/her/him/it’)

Loss of PTG object marking

Loss of PTG object marking in TG (J. 1987, 90, 98):

- Object marking still (partially) attested in Mbyá, Kaiwá, Chiriguano, Guarayu and Tupínamba
  - Diasporic TG languages in Michael et al.'s classification
- Object marking has disappeared in Wayampi, Urubú-Kaapor, Guajajára, Assuriní, Tapirapé, Kamaiurá, Parintintín and Kayabí
  - Non-diasporic TG languages in Michael et al.'s classification (except presumably Guajajára, cf. Tembé)
  - In Urubú-Kaapor and Guajajará, object marking has been reanalized as part of the root of some transitive verbs

Loss of PTG object marking



Michael et al. 2015

Other cross-referencing changes

Proto-Tupi-Guarani

	Indep	Ob-Top	Subord	Serial	Nominal
Non-ag S	2	2	2	2	2
Ag S	1	2	2	3	2
A>P	1-2,4	2	2	2	2
P>A	2	2	2	2	2

Wayampi

	Indep	Ob-Top	Subord	Serial	Nominal
Non-ag S	2	2	2	2	2
Ag S	1	1	1	1	2
A>P	1,4	1,4	1,4	2	2
P>A	2	2	2	2	2

Mbya Guarani

	Indep	Ob-Top	Subord	Serial	Nominal	
					CIRC	AG
Non-ag S	2	2	2	2	2	
Ag S	1	1	1	1	1	
A>P	1-2,4	1-2,4	1-2,4	2,4*	1-2	2
P>A	2	2	2	2	2	2

*refers to restricted set only

Kaiwa and Chiriguano

	Indep	Ob-Top	Subord	Serial	Nominal	
					CIRC	AG
Non-ag S	2	2	2	2	2	
Ag S	1	1	1	1	1	
A>P	1-2,4	1-2,4	1-2,4	1-2,4	1-2*	2
P>A	2	2	2	2	2	2

*4 also occurs in Kaiwa

Urubu

	Indep	Ob-Top	Subord	Serial	Nominal
Non-ag S	2	2	2	2	2
Ag S	1	1	1	1	1
A>P	1	1	1	1	1
P>A	1	1	1	1	1

Comparative cross-references systems (Jensen 1990)

Object marking in Mbyá

Object marking in Mbyá

Class 2 object marking disappeared following lenition **c* > **h* > ∅

Not all class 1 verbs have retained object marking:

(8) A-**i**-pota.
'I want them/her/him/it'

(9) A-poi.
'I released them/her/him/it'

Hypothesis:

- Loss of object marking is still in progress.
- Comparison of stems that have retained OM with those that lost allows us to elucidate the nature of this process.

Object marking in Mbyá

Class 1 object marking:

- allomorphs: *i-* before consonants, *j-* $\sim$ *nh-* before vowels
- is in complementary distribution with (de)transitivity markers and incorporation:
  - causative prefixes
  - reflexive prefix
  - reciprocal prefix
  - object incorporation

Object marking in Mbyá

jo- is no longer an object marker:

- Reflex of **jo-*, which Jensen takes to be an allomorph of **i-*
- *jo-* in Mbyá:
  - is attested with inactive cross-referencing of object:

    xe-joi, ‘They/she/he wash me’
  - co-occurs with transitivity (reduction) markers and object incorporation

    a-je-joi, ‘I wash myself’

Corpus study

Inventory of transitive stems

295 transitive stems were extracted from Dooley's (2016) lexicon of Mbyá

251 stems remained after exclusion of redundant variants

- e.g., *apo* 'do', *apo porã* 'do well', *apo vai* 'do poorly'

Each stem was coded for a series of 10 features.

Stem features 1

- Object marking: presence of object marking prefix on stem
- Loan: stem is a loan from Brazilian Portuguese
- Inflection class: class 1 versus class 2
- (De)-transitivization and incorporation:
  - jo-derived: stem derived by prefixation with *jo-*
  - mbo-derived: stem derived by causative prefix *mbo-*
  - ero-derived: stem derived by comitative prefix *ero-*

Stem features 2

- semantic features:
  - verb class, building on Levin (1993)
  - sense cluster: clustering of phrase-BERT embeddings of Dooley's definitions
- onset: phoneme used as onset of stem

  e.g., a-i-**p**ota
- frequency: frequency of stem in 245,516 words corpus

Categorical absence of object marking

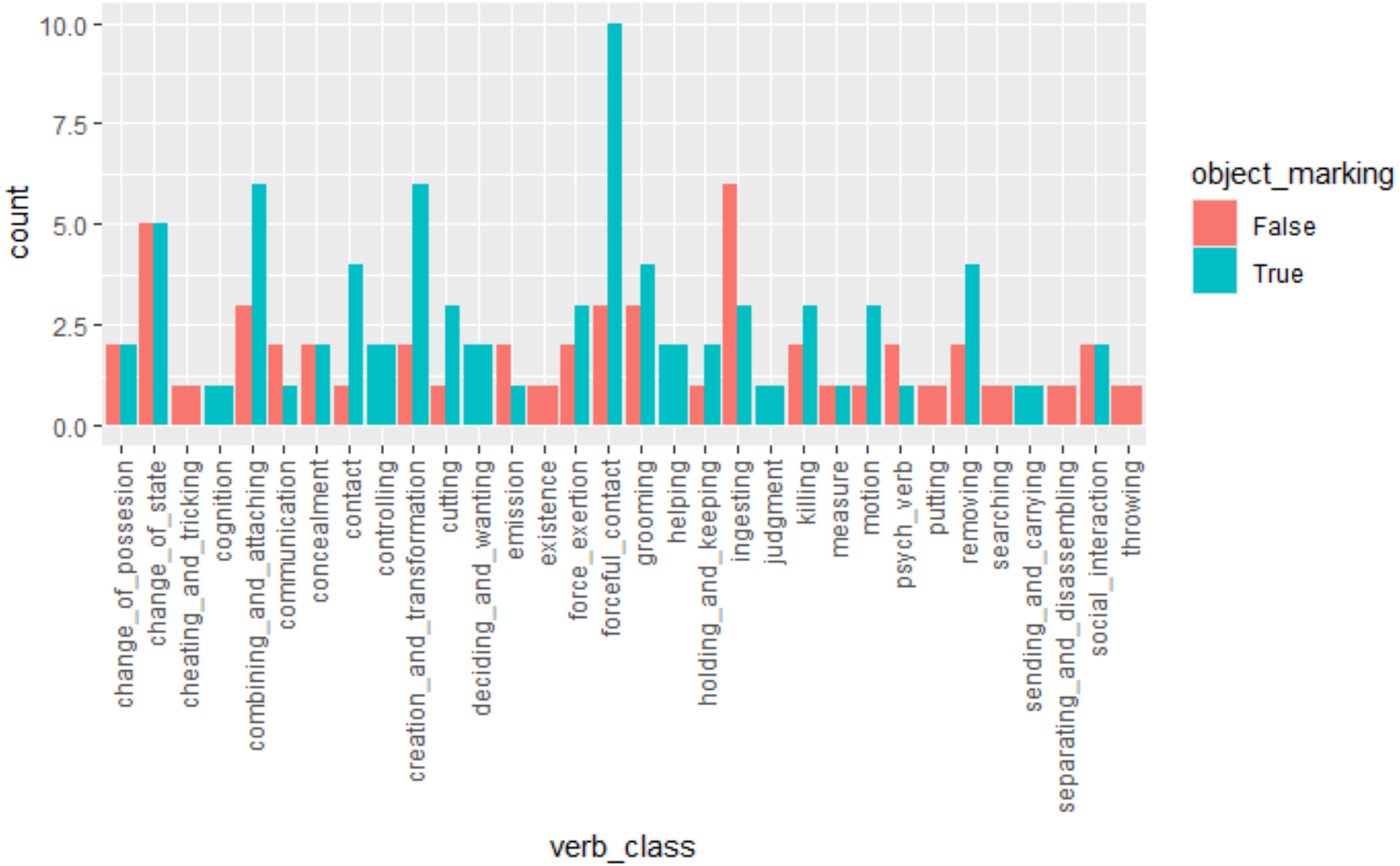
Some stem classes systematically lack object marking:

- Class 2 stems (e.g. *axa* ‘pass by, cross’)
- *jo*-derived stems (e.g. *joka* ‘break’)
- Causativized stems (e.g. *mbo’e* ‘teach’, *eru* ‘come’)
- B.P. loans (e.g. *gata* ‘spend’)

These stem classes were excluded from the study

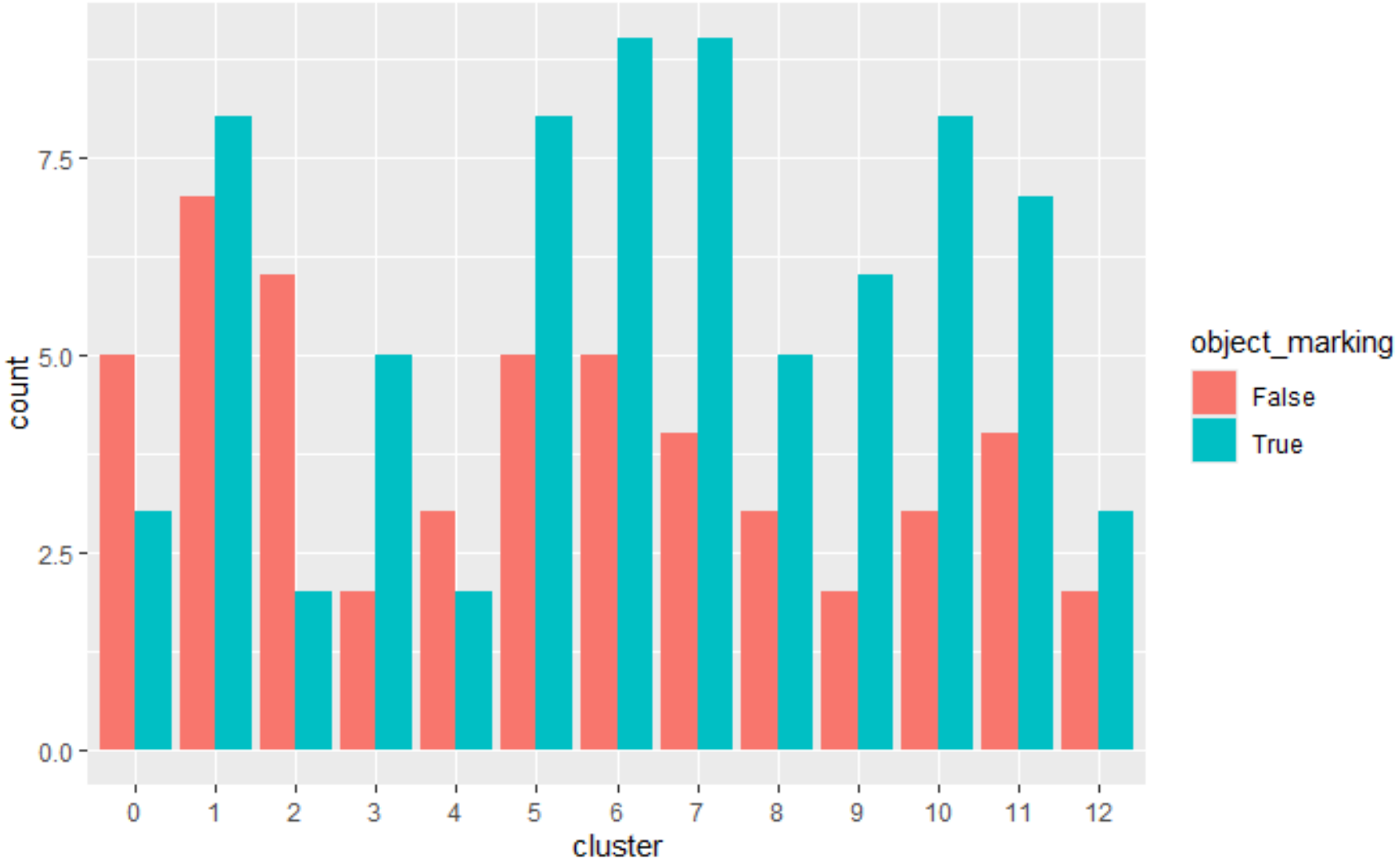
The remaining subset of stems includes 126 items

Object marking and stem sense: Levin verb classes



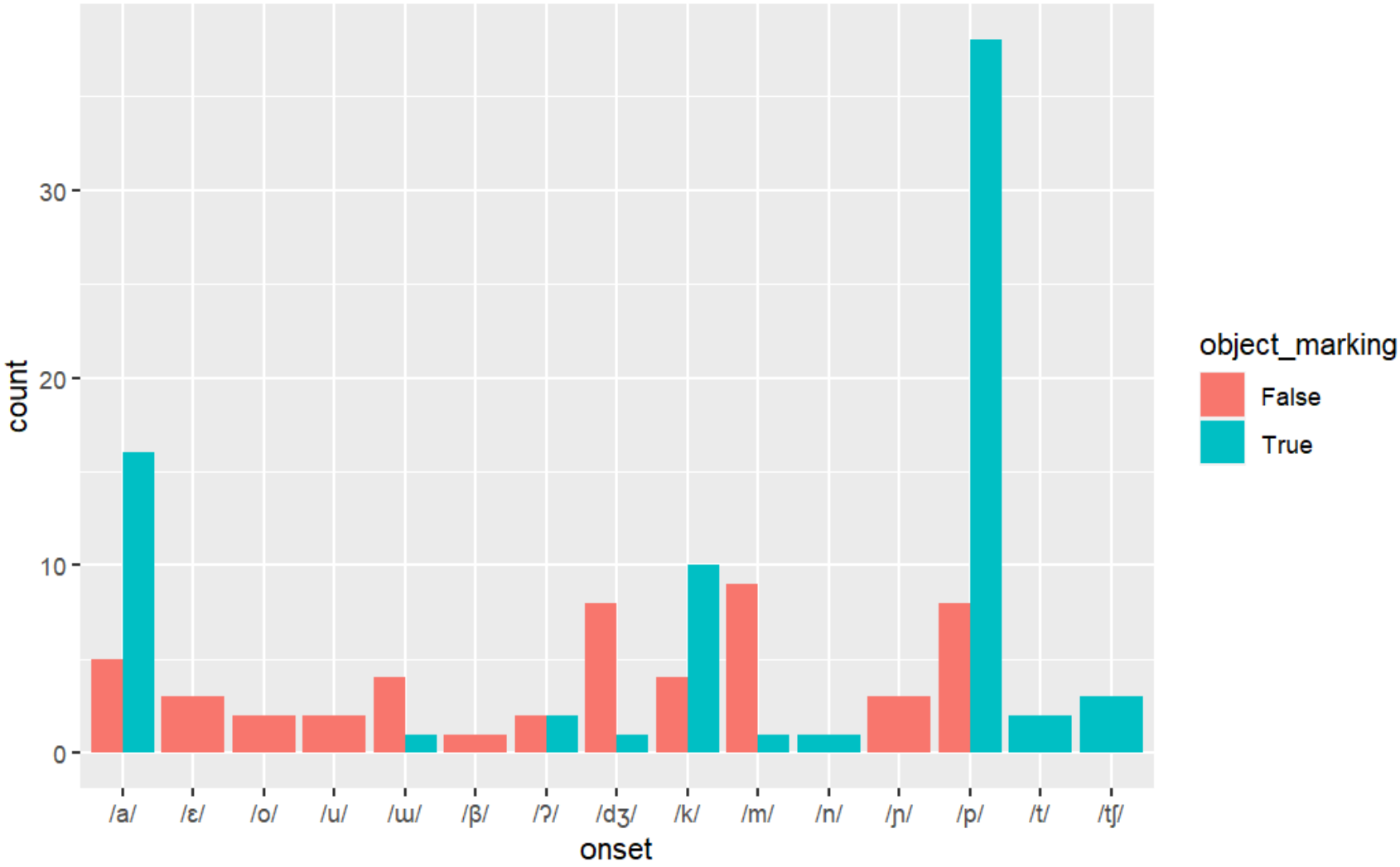
Fisher's Exact Test, p = 0.86

Object marking and stem sense: embeddings clusters



Fisher's Exact Test, p = 0.72

Object marking and stem onset



Fisher's Exact Test, $p < 0.001$

Object marking and phonological distance from onset

Hypothesis: the smaller the phonological distance between the object marker and the stem onset, the more likely it is for the object marker to be lost

Complication: vocalic allomorph of OM precedes consonantal onsets, consonantal allomorph precedes vocalic onset

Consequence: we must analyze consonant and vowels in the same feature space

Shared feature geometry for V and C place of articulation: Sagey (1986), McCarthy (1988), Clements (1991) and others

Object marking and phonological distance from onset

V and C shared features:

- Place of articulation (following Clements 1991):
  - Labial
  - Coronal
  - Dorsal
  - Radical
- Sonorant
- Voiced

Hamming distance: $\frac{\text{number of different-valued features}}{\text{total number of features}}$

Object marking and phonological distance from onset

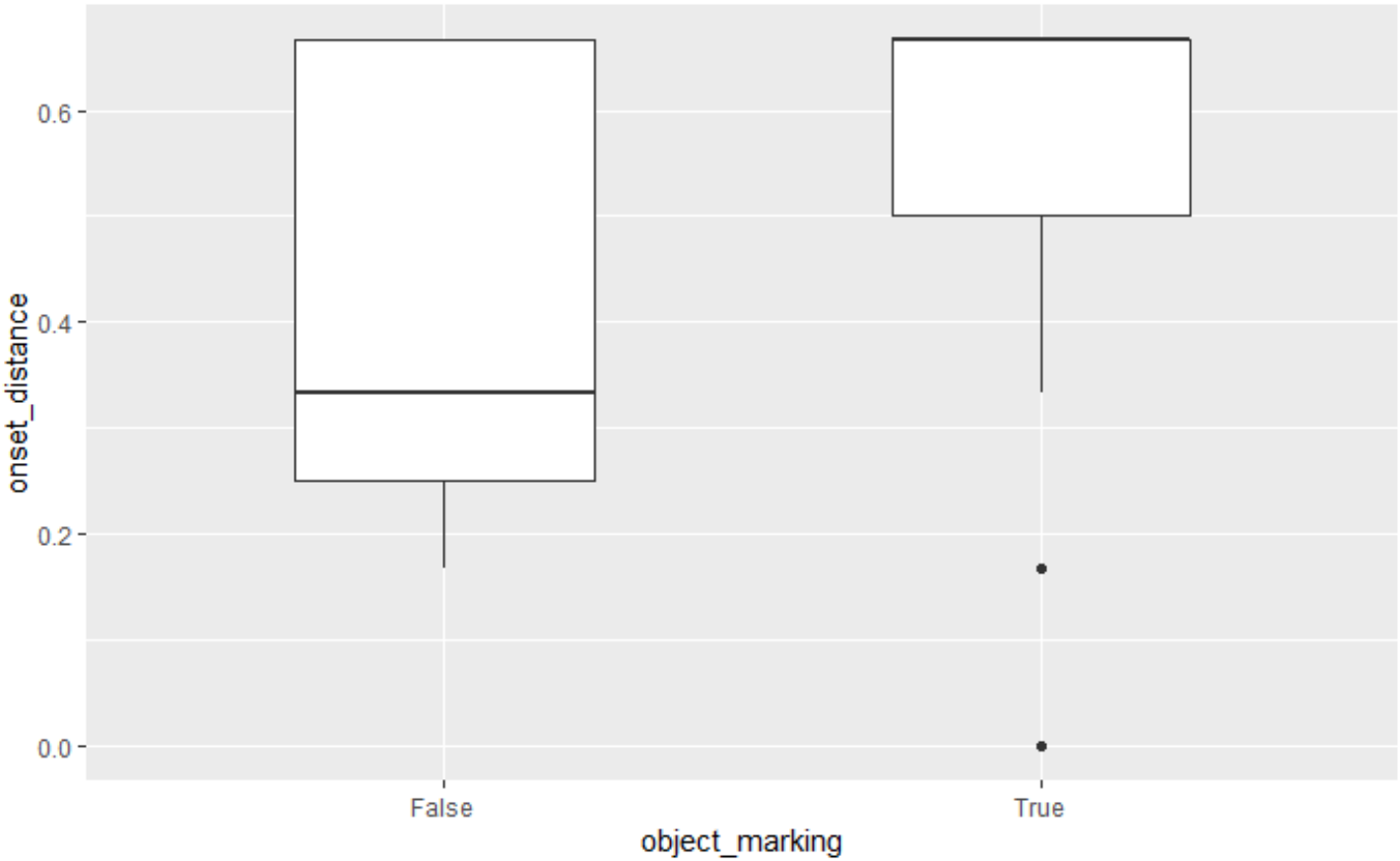
Clements' (1991) definition of V-place features:

- "labial characterizes vowels produced with a constriction at the lips (rounded vowels)"
- "coronal characterizes vowels produced with a constriction at the tip, blade or front of the tongue (front and retroflex vowels as opposed to central and back vowels)
- "dorsal characterizes vowels produced with a constriction of the center or back of the tongue, i.e. the palatine dorsum (back vowels as opposed to front and central vowels)"
- "radical characterizes vowels produced with a constriction in the lower pharynx (low and pharyngealized vowels)"

Object marking and phonological distance from onset

	radical	dorsal	coronal	labial	sonorant	voiced
a	1	0	0	0	1	1
ɛ	0	0	1	0	1	1
i	0	0	1	0	1	1
o	0	1	0	1	1	1
u	0	1	0	1	1	1
ɯ	0	1	0	0	1	1
β	0	0	0	1	1	1
dʒ	0	0	1	0	0	1
k	0	1	0	0	0	0
m	0	0	0	1	1	1
n	0	0	1	0	1	1
ɲ	0	1	1	0	1	1
p	0	0	0	1	0	0
t	0	0	1	0	0	0
tʃ	0	0	1	0	0	0

Object marking and phonological distance from onset



Wilcoxon Rank Sum Test, $p < 0.001$

Object marking and frequency

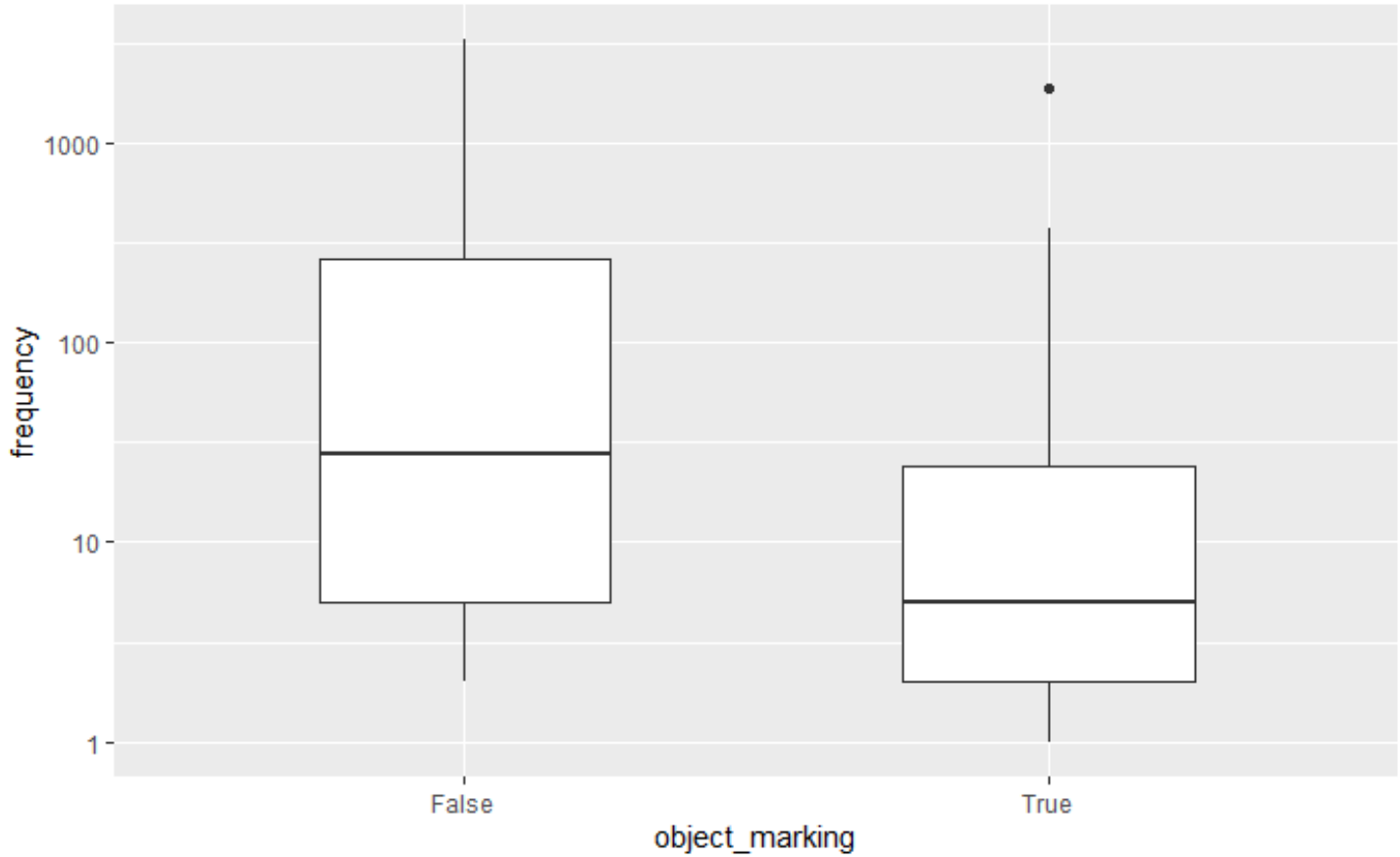
Corpus for stem frequency:

- Oratory discourse: 17,856 words
- Narratives: 12,150 words
- New testament: 214,618 words

Rank frequencies of transitive stems in biblical and non-biblical texts are strongly positively correlated:

- Kendall’s tau = 0.63 ($p < 0.001$)
- Biblical and non-biblical sub-corpora similar in relevant respects

Object marking and frequency



Wilcoxon Rank Sum Test, p = 0.06

Model of object marking loss

Stepwise logistic regression of object marking loss

Table: Selected Model

	Dependent variable:
	object_marking
Constant	-2.611^{***}
	(0.681)
frequency	-0.002^{**}
	(0.001)
onset_distance	6.159^{***}
	(1.272)
Observations	126
R^2	0.326
χ^2	34.884^{***} (df = 2)
Note:	*p<0.1; **p<0.05; ***p<0.01

Boostrap validation with 1,000 resamples: AUC-ROC = 0.8

Discussion

Loss of object marking as dissimilation

Loss of object marking in Mbyá is phonologically motivated:

- Dissimilation as consonant vowel interaction
- Deletion as dissimilation process

Loss of object marking as dissimilation

Consonant-vowel assimilation and dissimilation (Padgett 2011):

- Assimilation: non-low Vs in Cantonese must be front when between coronal consonants (Cheng 1991):
  - [tit] (‘iron’), [tøn] ‘shield’ but *[tut] or [ton]
- Dissimilation: Cantonese syllable rhyme cannot have both a rounded vowel and a labial coda (Cheng 1991):
  - *[up]

Loss of object marking as dissimilation

Deletion as dissimilation (Bye 2011):

- “Dissimilation is occasionally also used to refer to the deletion of one of a pair of similar neighboring sounds. Hall (2009), for example, describes this phenomenon with reference to /r/ in American English, in principle giving alternations like [fɑɹm] *farm* vs. [fɑmɚ] *farmer*, and [istɚn] *eastern* vs. [istənɚ] *easterner*.”

Loss of object marking as dissimilation

Features used in calculation of distance between object marker and stem onset:

- radical, dorsal, coronal, labial, sonorant, voice

Features attested to participate in dissimilation (Bye 2011):

- *labial*, *coronal*, lateral, rhotic, *voice*, spread glottis, *constricted glottis*, nasal, NC, continuant, high, low, length, H, L

One feature stands out (Bye 2011):

- “Major class features such as [consonantal], [sonorant], and [approximant] do not appear to participate in dissimilation.”
- Model AUC using [soronant] vs [continuant]: 0.8 vs 0.77

Frequency

- Infrequent stems are more resistant to object marking loss.
- This is the expected relation between sound change and word frequency (Hooper[Bybee] 1976):

> "(...) as Schuchardt observed in 1885, sound changes affect the most frequent lexical items first (1885 [1972]). And Paul, around 1886, observed that analogical leveling tends to affect infrequent items first (1890 [1972]). Or, stated differently, infrequent items are the most resistant to phonetically motivated change, while frequent items are the most resistant to conceptually motivated change."

Object marking loss vs other x-referencing changes

- Jensen (1998) on object marking:

  > “Its elimination does not appear to have any connection with other cross-referencing changes (since these have occurred in the Guaranian languages and Wayampi and Urubú-Kaapor).”

- Different processes:
  - Object marking loss is phonetically driven change
  - Other cross-referencing changes do not involve sound change so much as extension of active-inactive indexing and hierarchical indexing from independent to subordinate constructions

Object marking loss vs other x-referencing changes

Proto-Tupi-Guarani

	Indep	Ob-Top	Subord	Serial	Nominal
Non-ag S	2	2	2	2	2
Ag S	1	2	2	3	2
A>P	1-2,4	2	2	2	2
P>A	2	2	2	2	2

Wayampi

	Indep	Ob-Top	Subord	Serial	Nominal
Non-ag S	2	2	2	2	2
Ag S	1	1	1	1	2
A>P	1,4	1,4	1,4	2	2
P>A	2	2	2	2	2

Mbya Guarani

	Indep	Ob-Top	Subord	Serial	Nominal	
					CIRC	AG
Non-ag S	2	2	2	2	2	
Ag S	1	1	1	1	1	
A>P	1-2,4	1-2,4	1-2,4	2,4*	1-2	2
P>A	2	2	2	2	2	2

*refers to restricted set only

Kaiwa and Chiriguano

	Indep	Ob-Top	Subord	Serial	Nominal	
					CIRC	AG
Non-ag S	2	2	2	2	2	
Ag S	1	1	1	1	1	
A>P	1-2,4	1-2,4	1-2,4	1-2,4	1-2*	2
P>A	2	2	2	2	2	2

*4 also occurs in Kaiwa

Urubu

	Indep	Ob-Top	Subord	Serial	Nominal
Non-ag S	2	2	2	2	2
Ag S	1	1	1	1	1
A>P	1	1	1	1	1
P>A	1	1	1	1	1

Comparative cross-references systems (Jensen 1990)

References

Bye, Patrick. 2011. Dissimilation. In The Blackwell Companion to Phonology. Edited by Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume, and Keren Rice. John Wiley & Sons. Chapter 60.

Clements, G. N. 1991. Place of articulation in consonants and vowels: A unified theory. Working Papers of the Cornell Phonetics Laboratory 5. 77–123.

Hooper, Joan B. 1976. Word frequency in lexical diffusion and the source of morphophonological change. In Current progress in historical linguistics, ed. William Christie, 95–105

Jensen, Cheryl. 1987. Object prefix incorporation in Proto Tupí-Guaraní verbs. Language Sciences 9:45-55.

Jensen, Cheryl. 1990. Cross-referencing changes in some Tupi-Guarani languages. Amazonian Linguistics: Studies in Lowland South American Languages, ed. Doris L. Payne, pp. 117-158. Austin: University of Texas Press.

Jensen, Cheryl. 1998. "Comparative Tupí-Guaraní morphosyntax." In Desmond C. Derbyshire and Geoffrey K. Pullum (orgs.), Handbook of Amazonian languages, vol. 4, Berlin, Mouton de Gruyter, 489-618.

Levin, Beth. 1993. English verb classes and alternations. The University of Chicago Press.

McCarthy, John J. 1988. Feature geometry and dependency: A review. Phonetica 45. 84–108.

Michael, Lev, Natalia Chousou-Polydouri, Keith Bartolomei, Erin Donnelly, Vivian Wauters, Sérgio Meira and Zachary O'Hagan. 2015. A Bayesian Phylogenetic Classification of Tupí-Guaraní. Llames 15(2): 193-221.

Padgett, Jaye. 2011. Consonant-Vowel Place Feature Interactions. In The Blackwell Companion to Phonology. Edited by Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume, and Keren Rice. John Wiley & Sons. Chapter 75.

Sagey, Elizabeth. 1986. The representation of features and relations in nonlinear phonology. Ph.D. dissertation, MIT.