

Universal Dependencies for Mbyá Guaraní

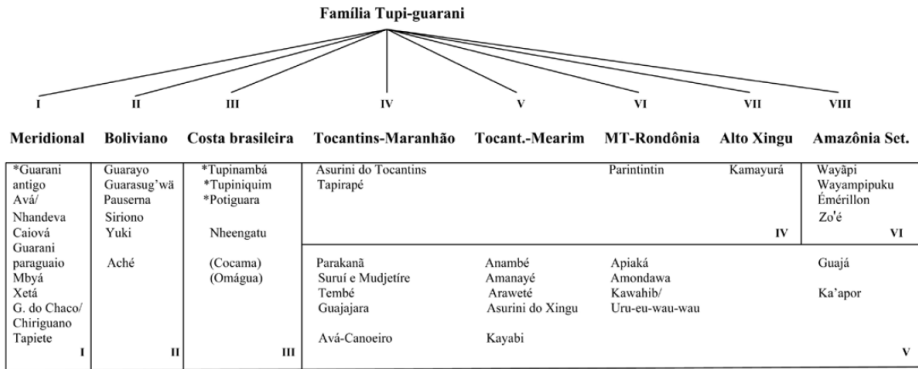
Guillaume Thomas

August 30, 2019

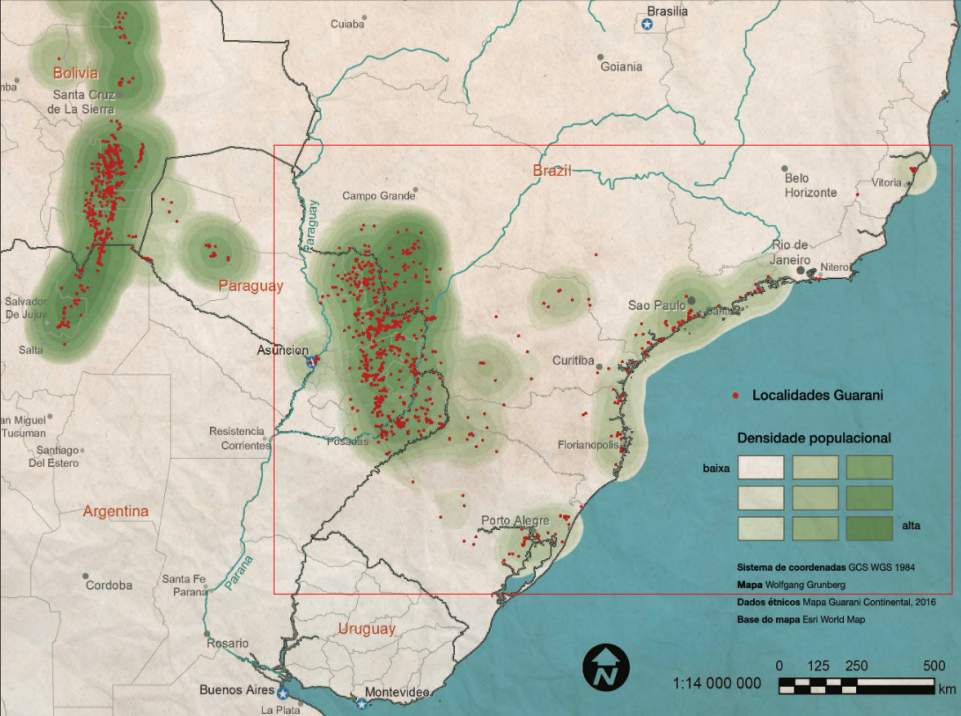
Department of Linguistics
University of Toronto

Mbyá Guaraní

- Tupi-Guaraní language
- About 30,000 speakers: Argentina, Brazil, Paraguay



(Dietrich 2010)



Corpus

- UD Mbyá Guaraní Dooley:

Robert A. Dooley. 2011 Mbyá Guaraní collection of Robert Dooley. The Archive of the Indigenous Languages of Latin America: www.ailla.utexas.org. Media: text. Access: 100% restricted. PID ailla:119734.

Guillaume, Thomas and Dooley, Robert A. 2019. Dependency Treebank derived from the Mbyá Guaraní collection of Robert Dooley. Access: 100% restricted. PID ailla:119734

- 33 narratives, 1046 sentences
- 2 authors, Rio das Cobras, Paraná, Brazil

- UD Mbyá Guaraní Thomas

- Tiny 98 sentence corpus of autobiographical narratives recorded in Paraguay

Corpus

- Modification to Dooley's interlinearization in SIL FLEx
- Features converted from FLEx glosses and tags
- Dependency annotation:
 - manual annotation of first 500 sentences in Arborator
 - UDPipe annotation of second half, manual correction
 - first round of correction, four student RAs
 - three other rounds by PI

This talk

- Properties of Mbyá that challenge current UD annotation scheme
- Favour alternatives already suggested in earlier work:

Kim Gerdes, Sylvain Kahane. 2016. Dependency Annotation Choices: Assessing Theoretical and Practical Issues of Universal Dependencies. In proceedings of LAW 10, ACL, 131–140.

William Croft, Dawn Nordquist, Katherine Looney, Michael Regan. 2017. Linguistic Typology meets Universal Dependencies. In proceedings of TLT15, 63–75.

Kim Gerdes, Bruno Guillaume, Sylvain Kahane, Guy Perrier. 2018. SUD or Surface-Syntactic UniversalDependencies: An annotation scheme near-isomorphic to UD. In proceedings of UDW 2018, 66–74.

Syntactic Categories

Nouns and Verbs

- Morphology: nouns morphologically similar to inactive verbs
- Syntax:
 - nouns are productively predicative
 - predicative nouns behave as a mixed category
- Matter of debate among Guaranologists:

Wolf Dietrich. 2017. Word Classes and Word Class Switching in Guaraní Syntax. In Bruno Estigarribia and Justin Pinta (eds), *Guaraní Linguistics in the 21 st century*, pages 158–193. Leiden: Brill.

Nouns and Verbs

- Noun:
 - can be used as argument without derivation
 - compatible with nominal tense

(1) A-japo xe-r-o-rã.

A3-do B1.SG-R-house-FUT

VERB NOUN

vt n

'I am building my house.' (Dooley 2015)

- Note form of possessive prefix

Nouns and Verbs

- Active/inactive alignment:

(2) A-vaẽ.

A1.SG-arrive

VERB

vi:a

'I arrived.'

(3) Xe-kane'õ.

B1.SG-tired

VERB

vi:i

'I am tired.'

- Inactive verbs and nouns belong to the same agreement inflection class

Nouns and Verbs

- Predicative uses of nouns:

(4) Xe-irũ.

B1.SG-friend

'I have a friend.'

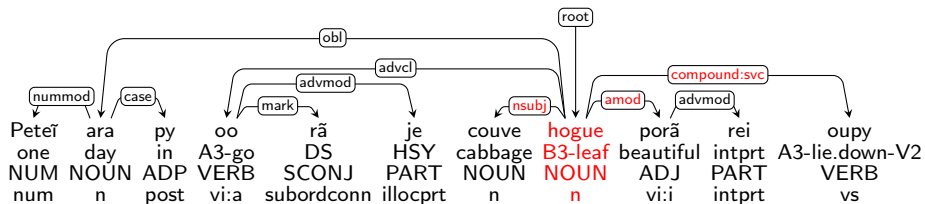
(5) João xe-irũ.

João B1.SG-friend

'João is my friend.'

Nouns and Verbs

- Predicative nouns as a mixed category:



'One day, when he went [there], [he saw that] the cabbage had beautiful leaves.'

- Tagged as NOUN:
 - Analyzed as predicate nominal constructions
 - Other languages may use copular/verbal strategies for this construction
 - cf. Croft et al. (2017)

Adjectives and Adverbs

- No morphological categories of adjectives and adverbs
- Stative verbs used as modifiers:

(6)	Kova'e	ára	ma	i-porã	vaipa.
	DEM	day	BDY	B3-good	very
	DET	NOUN	PART	VERB	PART
	dem	n	discprt	vi:i	intprt

'This day is very good.' (Dooley 2015)

- Here categorization favours syntactic rather than morphological information

Adjectives and Adverbs

- No morphological categories of adjectives and adverbs
- Stative verbs used as modifiers:

(7)	Avaxi	o-nhotỹ	r-yxy	porã.
	Corn	A3-plant	R-line	good
	NOUN	VERB	NOUN	ADJ
	n	vt	n	vi:i

'He planted the corn in beautiful lines.' (Dooley 2015)

- Here categorization favours syntactic rather than morphological information

Adjectives and Adverbs

- No morphological categories of adjectives and adverbs
- Stative verbs used as modifiers:

(8) Oro-vy'a	porã.
A1.PL.EXCL-happy	good
VERB	ADV
vi:a	vi:i

'We were very happy.' (Dooley 2015)

- Here categorization favours syntactic rather than morphological information

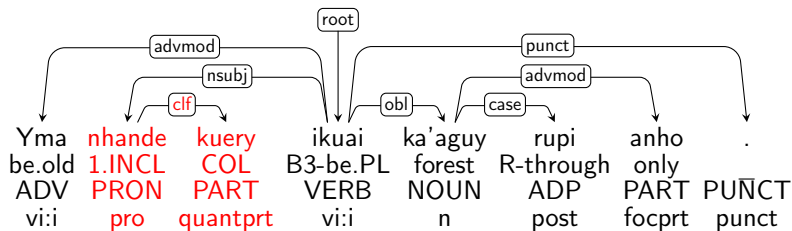
Dependencies

Particles

- Uninflected
- Short (one or two syllables)
- Flexible with respect to the category of their head
- Functions:
 - Express grammatical features of their head (e.g. aspect)
 - Non-determiner quantifiers
 - Focus sensitive operators
 - Illocutionary modifiers

Issues with nominal particles

- Do not match any UD nominal dependent
- Example: collective/associative plural particle *kuery*

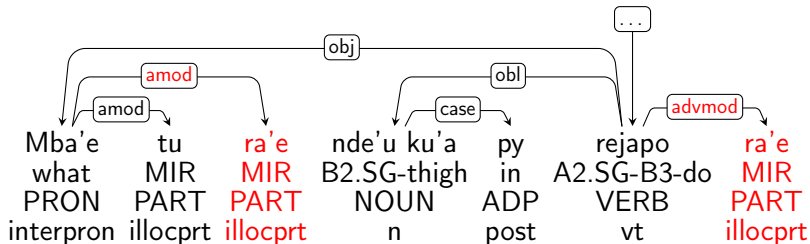


'A long time ago, we lived in the forest.'

- Unsatisfying decision: *kuery* introduced by *clf*

Issues with TAME particles

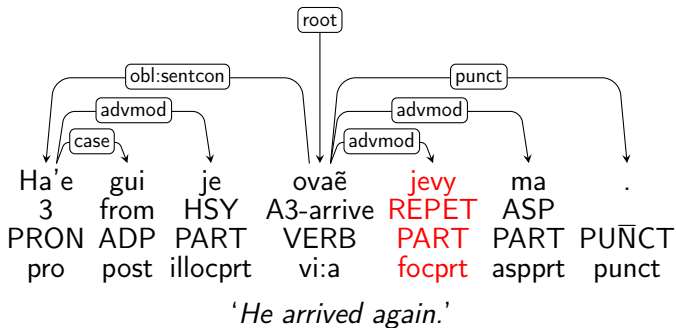
- Reluctant to relate them to their head by aux
- Modification of nouns as well as verbs:



'What did you do to your thigh?'

Issues with TAME particles

- Reluctant to relate them to their head by aux
- TAME notions conveyed through adverbs in English:



Dependencies for particles

- Current annotation scheme (simplified):
 - Associative plural related to NOUN, PRON or PROPN by `c1f`
 - Interrogative particle *pa* introduced by `discourse:q`
 - Other PART related to NOUN/PRON/PROPN by `amod`
 - Other PART related to their heads by `advmod`
- A better solution: category neutral `mod` (Gerdes et al. 2018)

Particles

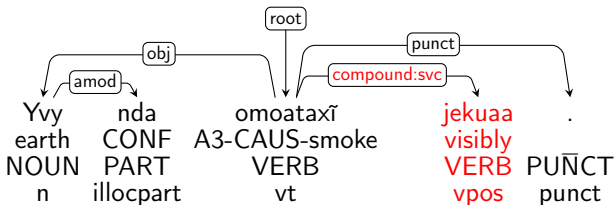
- Subcategorization of particles in language specific tagset makes it easy to change the label of these relations:

aspect particles	aspprt
discourse particles	discprt
focus particles	focprt
illocutionary particles	illocprt
intensifiers	intprt
modal particles	modprt
quantificational particles	quantprt
question particles	qprt
tense particles	tempprt

- e.g. map advmod to aux for aspprt modifiers of VERB

Postposed roots as compound:svc

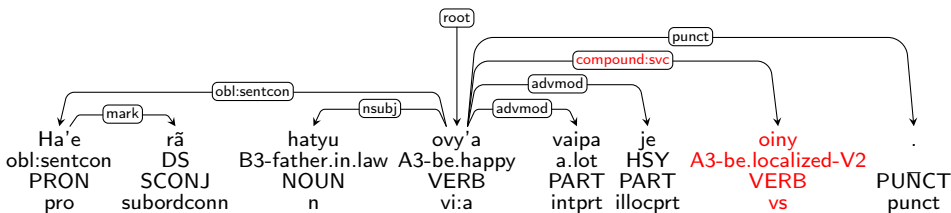
- share arguments and TAME
- uninflected
- no independent argument
- no argument or modifier intervening between verb and postposed root



'He even raised dust.'

Secondary predicates as compound:svc

- share arguments and TAME
- identified by a converbial suffix
- inflected for agreement in person and number
- some arguments or modifiers may intervene between predicates



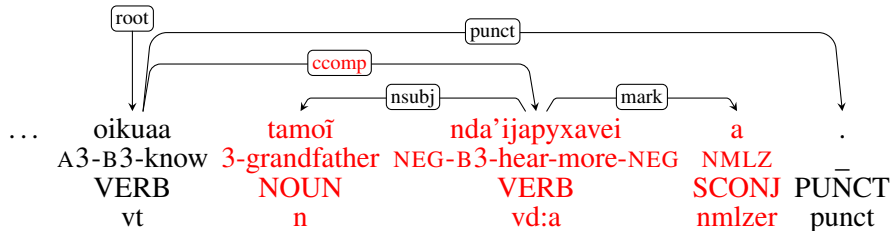
'And then his father in law rejoiced..'

Serial Verb Constructions as compound?

- 'Secondary predicates' don't show the level of morphological integration one would expect of compounds
- No satisfying alternative in current inventory of dependency relation labels
- Serial Verb Constructions are arguably forms of cosubordination (Olson 1981, Foley & Van Valin 1984):
 - more syntactic structure than compounds
 - neither coordination nor subordination
- A better solution? Croft et al. (2017) suggested *cxp*

Clausal nominalization as ccomp and csubj

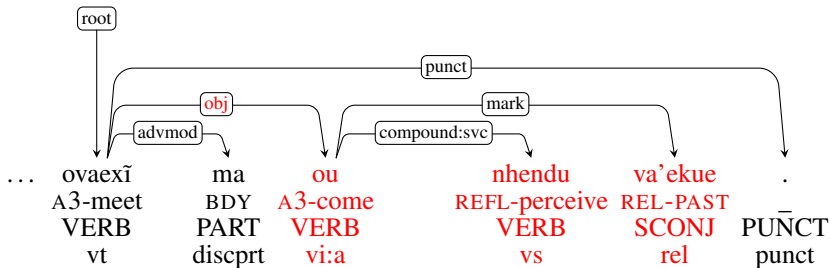
- Clausal properties:
 - internal clausal structure
 - denote propositions
- Nominal properties:
 - compatible with nominal tense suffixes
 - can be complement of postpositions



'He knew that his grandfather couldn't hear well anymore.'

Free relative clauses as nsubj and obj

- Clausal properties:
 - internal clausal structure
- Nominal properties:
 - denote entities
 - compatible with nominal tense suffixes
 - can be complement of postpositions



'He met the person that he had heard coming.'

Dependencies for nominalized clauses

- We are forced into a somewhat arbitrary choice:
 - `obj/nsubj`: resolve mixed category to NOUN
 - `ccomp/csubj`: resolve mixed category to a clause
- Better alternative (Croft et al. 2017, Gerdes et al. 2018)
 - `subj`
 - `comp`

Conclusion

Conclusion

- Phenomena at issue:
 1. Mixed categories (predicate nominals, nominalization)
 2. Category neutral modification (particles)
 3. Cosubordination (serial verb constructions)
- Issues arise at level of dependency relation labelling.
- Issue with 1 and 2: mixing POS and relation label (Gerdes et al. 2016, 2018; Croft et al. 2017)

Worry: if the mixing of POS and relation label leads to arbitrary annotation decisions, does it lead to less homogeneous annotation guidelines across languages?
- Issue with 3: need to add a new class of dependency relations besides coordination and subordination.

Thank You