Word order variation in Mbyá Guaraní

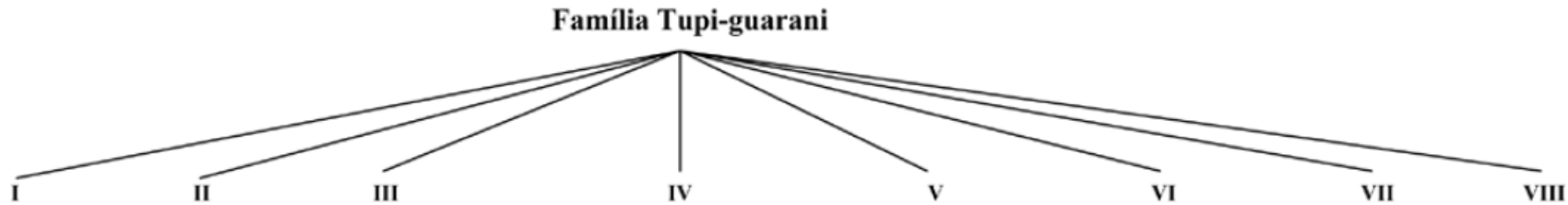
Angelika Kiss
Guillaume Thomas

August 30, 2019

Department of Linguistics
University of Toronto

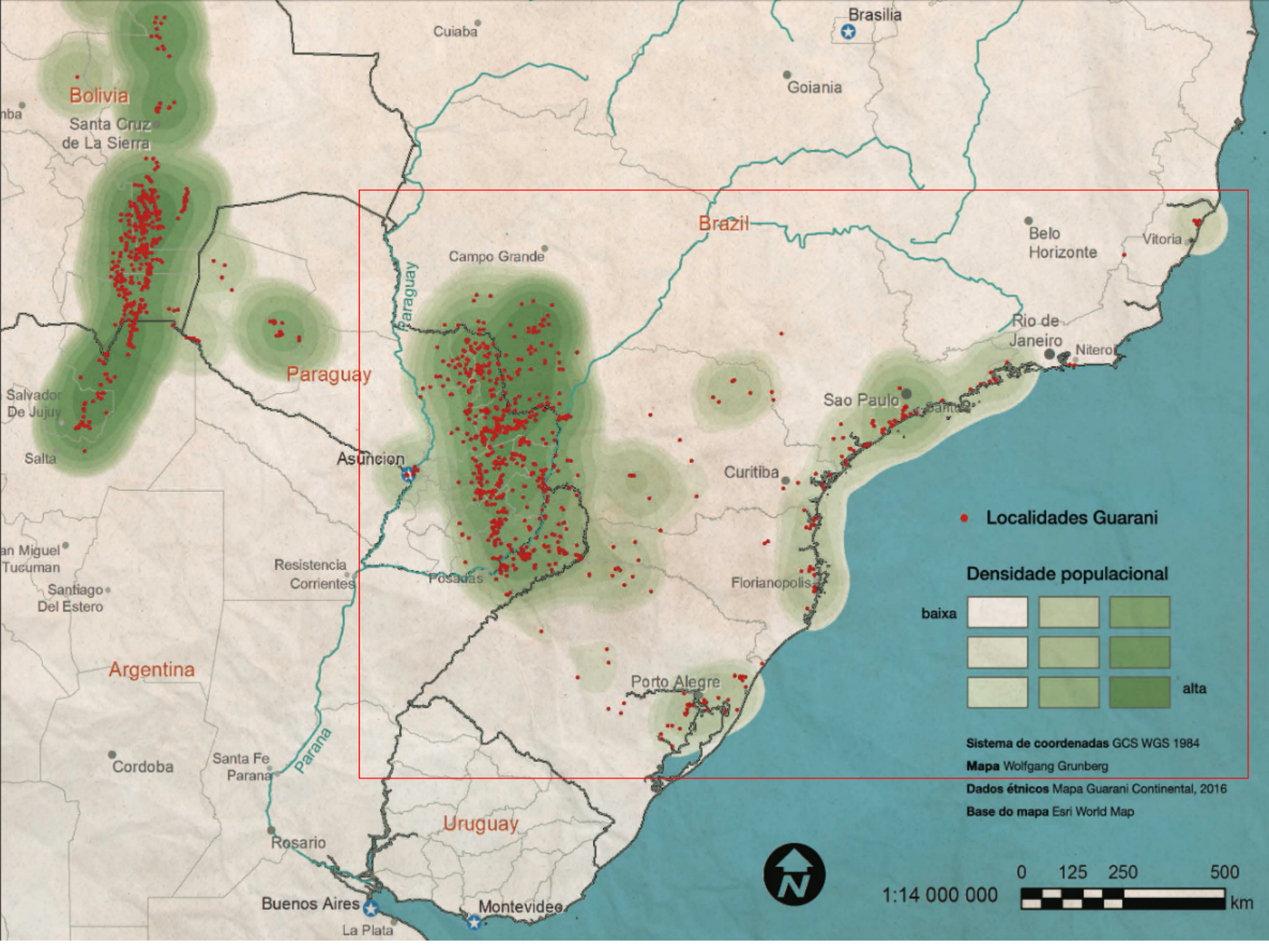
Word Order in Mbyá

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



Meridional	Boliviano	Costa brasileira	Tocantins-Maranhão	Tocant.-Mearim	MT-Rondônia	Alto Xingu	Amazônia Set.
*Guarani antigo Avá/ Nhandeva Caiová Guarani paraguaio Mbyá Xetá G. do Chaco/ Chiriguano Tapiete I	Guarayo Guarasug'wä Pauserna Siriono Yuki Aché II	*Tupinambá *Tupiniquim *Potiguara Nheengatu (Cocama) (Omágua) III	Asurini do Tocantins Tapirapé IV Parakanã Suruí e Mudjetíre Tembé Guajajara Avá-Canoeiro	Anambé Amanayé Araweté Asurini do Xingu Kayabi	Parintintin Apiaká Amondawa Kawahib/ Uru-eu-wau-wau	Kamayurá	Wayãpi Wayampipuku Émérillon Zo'é VI Guajá Ka'apor V

(Dietrich 2010)


Brasilia
Cuiaba
Goiania
Bolivia
Santa Cruz de La Sierra
Brazil
Belo Horizonte
Vitoria
Campo Grande
Paraguay
Rio de Janeiro
Niteroi
Paraguay
Salvador De Jujuy
Sao Paulo
Salta
Asuncion
Curitiba
Resistencia
Corrientes
Posadas
Florianopolis
Santiago Del Estero
Argentina
Porto Alegre
Cordoba
Santa Fe
Parana
Parana
Uruguay
Rosario
Buenos Aires
La Plata
Montevideo
Localidades Guarani
Densidade populacional
baixa
alta
Sistema de coordenadas GCS WGS 1984
Mapa Wolfgang Grunberg
Dados étnicos Mapa Guarani Continental, 2016
Base do mapa Esri World Map
N
1:14 000 000
0 125 250 500
km

Motivations

- Previous studies
    - Dooley (1985, 2015)
    - Martins (2003)
- Methodological issue/typological implications
    - Split-S (active/inactive) language
    - How should we describe core argument position?
    - S and O or A and P?

Grammatical background

Active/inactive alignment

Active/inactive intransitive verbs

(1) Xee a- a ju ma.
I A1.sg- go again already

'*I am already going again*.'

(2) Xe- kangy vaipa.
B1.SG- feel_weak very

'*I feel very weak*.'

Active/inactive alignment

Person hierarchy: 1 > 2 > 3

(3) A- exa.
A1.SG- R-

'*I saw him/her/it/them*.'

(4) Xe- r- exa.
B1.SG- R- see

'*They/(s)he/you saw me*.'

Grammatical functions

- Subject:
  - Unique cross-referenced argument of intransitive verb
  - Active argument of transitive verb
- Object:
  - Inactive argument of transitive verb

Velazquez Castillo (2002): no S and O in Guaraní

- Noun-incorporation targets non-actors (rather than objects)

(5) (Che) che- r- esa+ r- ovy.
I B1.SG- R- eye R- blue
'I am blue eyed.'

- Reflexivization is controlled by actor (rather than subject)

(6) Vierne santo n- o- ñe- mba'apó -i
Friday saint NEG- A3- REFL- work NEG
'On Good Friday one does not work.'

Velazquez Castillo (2002): no S and O in Guaraní

- Verb serialization does not mix actor/non-actor

(7) O- pu'ã o- guata.
A3- get up A3 walk
'He got up and walked.'

(8) *O- pu'ã i- mandu'a.
A3- get up B3 remember
'He got up and remembered.'

- Relativization gaps are not restricted by grammatical function

Dooley (2015): evidence for S and O in Mbyá

- Word order: S preverbal, O postverbal
- Reflexive voice is controlled by S
- Impersonal voice targets S

(9) O- u -a.
A3- come IMPR
'Someone came.'

(10) O- juka -a.
A3- kill IMPR
'Someone killed him/her.'

Dooley (2015): evidence for S and O in Mbyá

- Pivots in switch reference are S

(11) Ava o- exa mboi o- o vy
man A3 see snake A3 come SS
'The man$_1$ saw the snake$_2$ when he$_1$ came.'

- *embi-* and *-py* nominalizations denote objects

(12) xe- r- embi- exa
B1.SG- R OBJ_NMLZ see
'what I see'

(13) o- exa -py
A3- kill OBJ_NMLZ.SUBJ_IMPR
'what is seen'

This talk

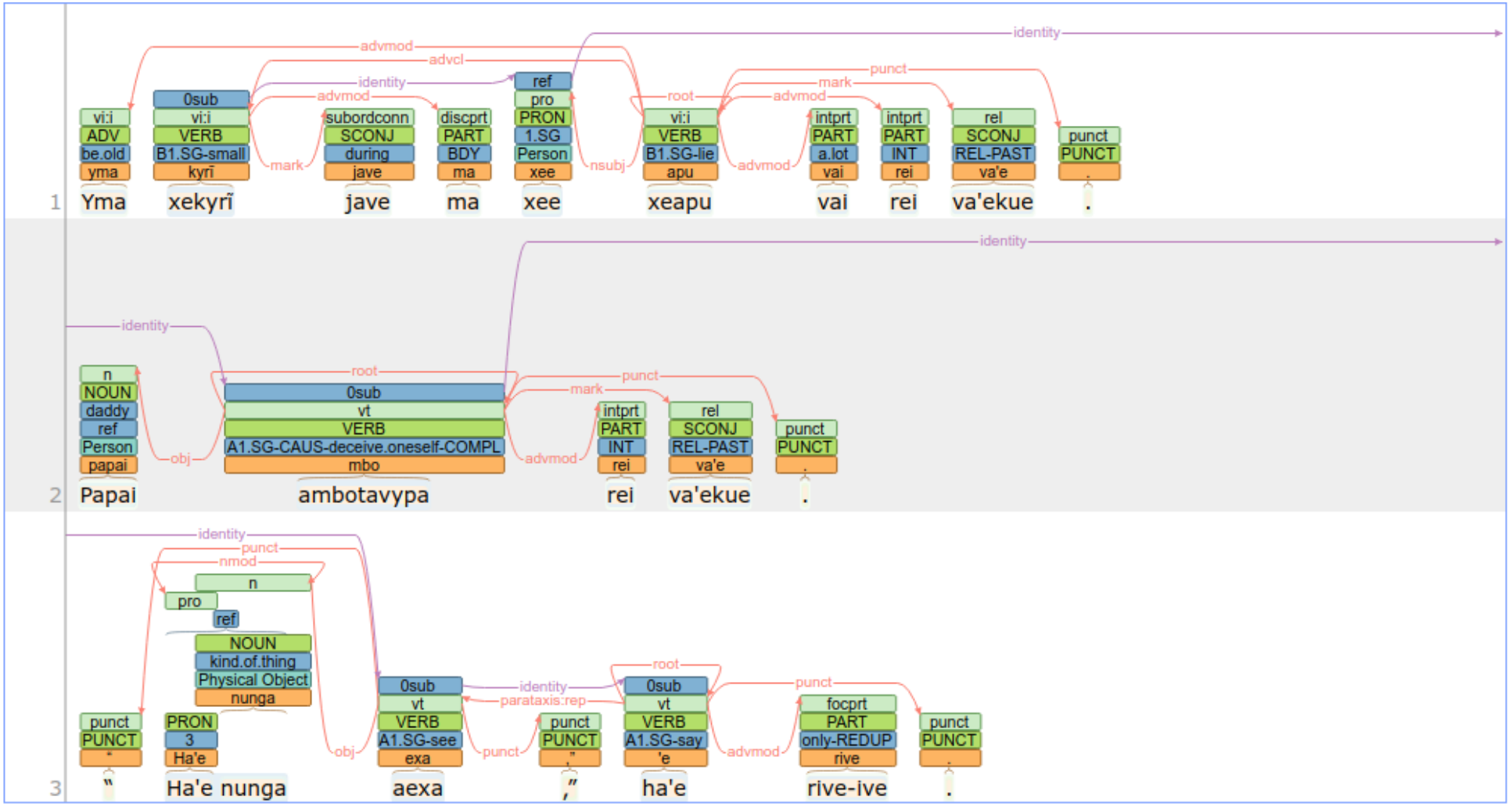
- Compare descriptions of word order by A/P vs S/O:
    - Do we miss generalizations with either option?
- Compare models of argument placement with A/P vs S/O as predictor:
    - How accurate is each model?
    - Do we miss interesting interactions by excluding either predictor?

Corpus and annotation layers

Corpus

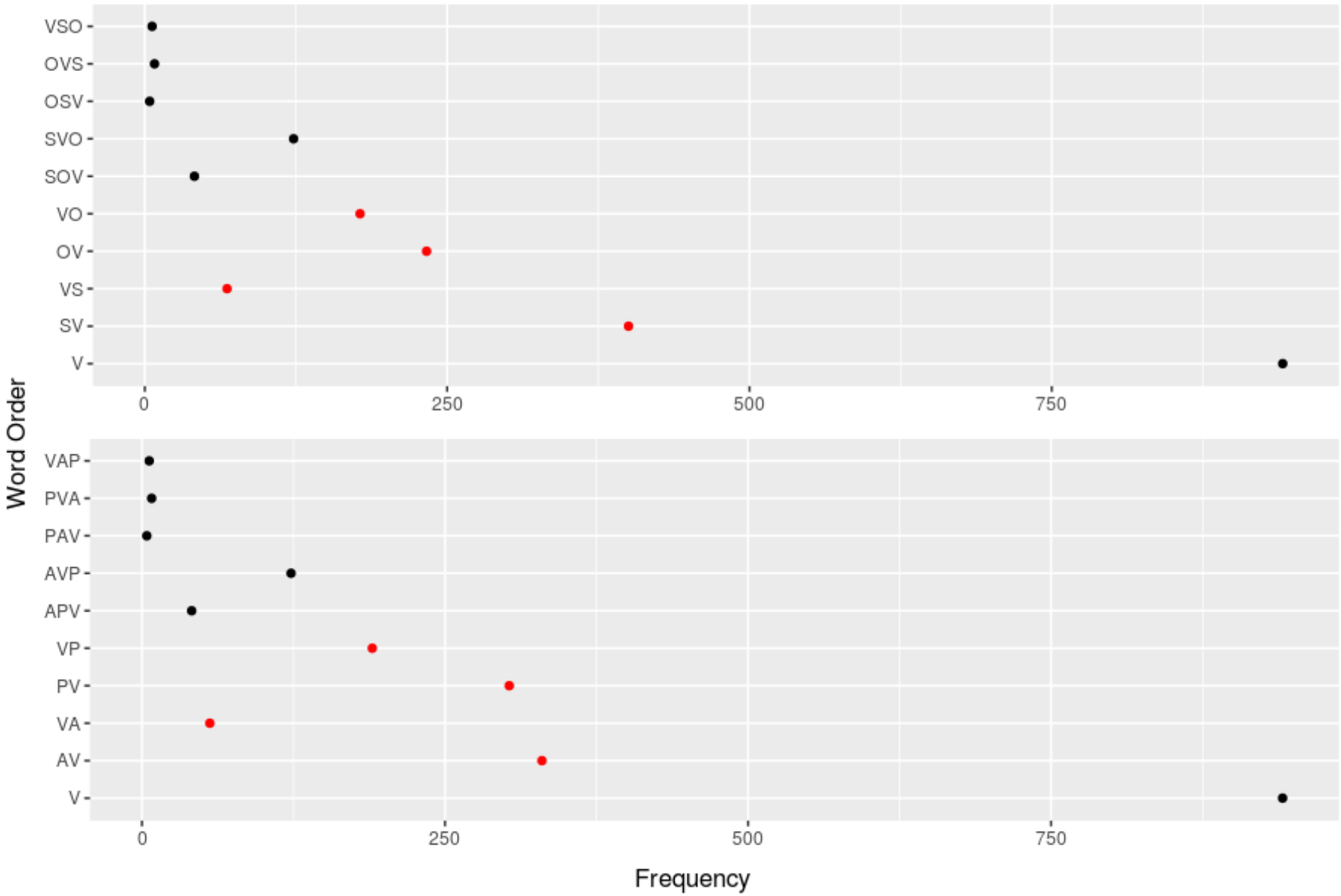
- Dooley's (2011) AILLA corpus:
  - 33 narratives, 1046 sentences
  - 2 authors, Rio das Cobras, Paraná, Brazil
- Interlinearization in SIL FLEx
- Dependency annotation in Arborator
- Coreference, ontological class annotation in Webanno3
- UD annotation available in UD v2.4

Annotation layers



Descriptive statistics

Word Order Overview



Argument Position

- Argument placement: preverbal (XV), postverbal (VX)
- Predictors:
  - Alignment: active, inactive
  - Animacy: animate, inanimate
  - Clause Type: root, subordinate
  - Givenness: given, new
  - Grammatical Function: subject (S), object (O)
  - Length: # characters in phrase
  - Transitivity: intransitive (vi), transitive (vt)

Argument position

		XV	VX	p
Alignment	active	$498_{88.0}$	$68_{12.0}$	<0.001
	inactive	223_{59}	$155_{41.1}$	
Animacy	animate	$578_{82.7}$	$121_{17.3}$	<0.001
	inanimate	$143_{58.4}$	$102_{41.6}$	
Clause Type	root	$568_{73.9}$	$201_{26.1}$	<0.001
	sub	$153_{87.4}$	$22_{12.6}$	
Givenness	given	$598_{81.8}$	$133_{18.2}$	<0.001
	new	$123_{57.7}$	$90_{42.3}$	
G. Function	S	$568_{88.1}$	$77_{11.9}$	<0.001
	O	$153_{51.2}$	$146_{48.8}$	
Length	Mean (SD)	$7.7_{4.1}$	$9.4_{4.1}$	<0.001
Transitivity	vi	$327_{85.2}$	$57_{14.8}$	<0.001
	vt	$394_{70.4}$	$166_{29.6}$	

Argument position by grammatical function

		Subjects			Objects		
		XV	VX	p	XV	VX	p
Animacy	animate	$533_{88.8}$	$67_{11.2}$	*	$45_{45.5}$	$54_{54.5}$	
	inanimate	$35_{77.8}$	$10_{22.2}$		$108_{54.0}$	$92_{46.0}$	
Clause Type	root	$461_{86.8}$	$70_{13.2}$	*	$107_{45.0}$	$131_{55.0}$	***
	sub	$107_{93.9}$	$7_{6.1}$		$46_{75.4}$	$15_{24.6}$	
Givenness	given	$510_{91.1}$	$50_{8.9}$	***	$88_{51.5}$	$83_{48.5}$	
	new	$58_{68.2}$	$27_{31.8}$		$65_{50.8}$	$63_{49.2}$	
Length	Mean	7.2	9.1	***	9.4	9.5	
Transitivity	vi	$327_{85.2}$	$57_{14.8}$	**			
	vt	$241_{92.3}$	$20_{7.7}$				

Argument position by alignment

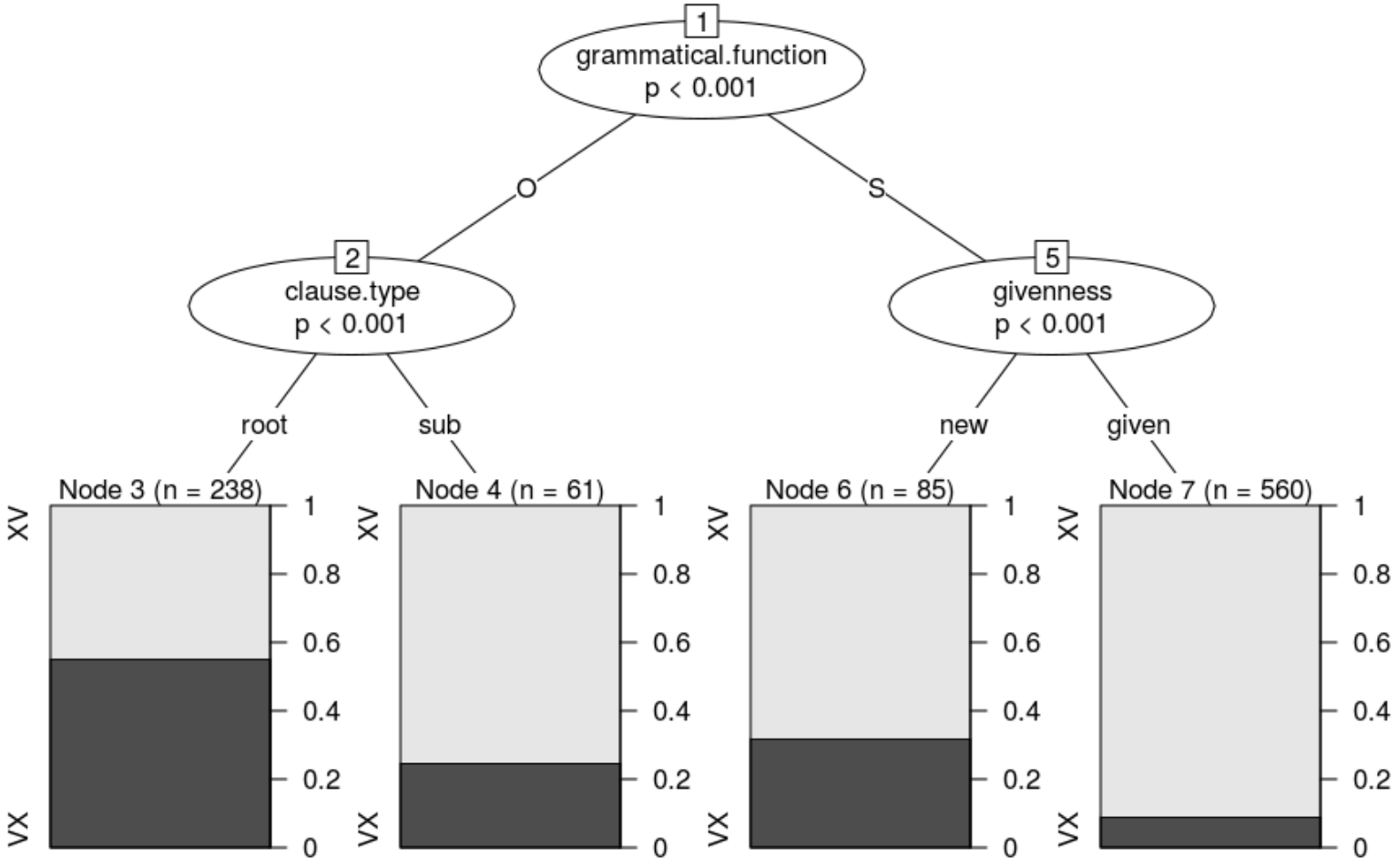
		Active			Inactive		
		XV	VX	p	XV	VX	p
Animacy	animate	$480_{88.2}$	$64_{11.8}$		$98_{63.2}$	$57_{36.8}$	
	inanimate	$18_{81.8}$	$4_{18.2}$		$125_{56.1}$	$98_{43.9}$	
Clause Type	root	$418_{87.1}$	$62_{12.9}$		$150_{51.9}$	$139_{48.1}$	***
	sub	$80_{93.0}$	$6_{7.0}$		$73_{82.0}$	$16_{18.0}$	
Givenness	given	$461_{91.3}$	$44_{8.7}$	***	$137_{60.6}$	$89_{39.4}$	
	new	$37_{60.7}$	$24_{39.3}$		$86_{56.6}$	$66_{43.4}$	
Length	Mean	7.1	9.2	***	8.9	9.5	*
Transitivity	vi	$257_{84.3}$	$48_{15.7}$	**	$70_{88.6}$	$9_{11.4}$	***
	vt	$241_{92.3}$	$20_{7.7}$		$153_{51.2}$	$146_{48.4}$	

Models of argument position

Models of argument position

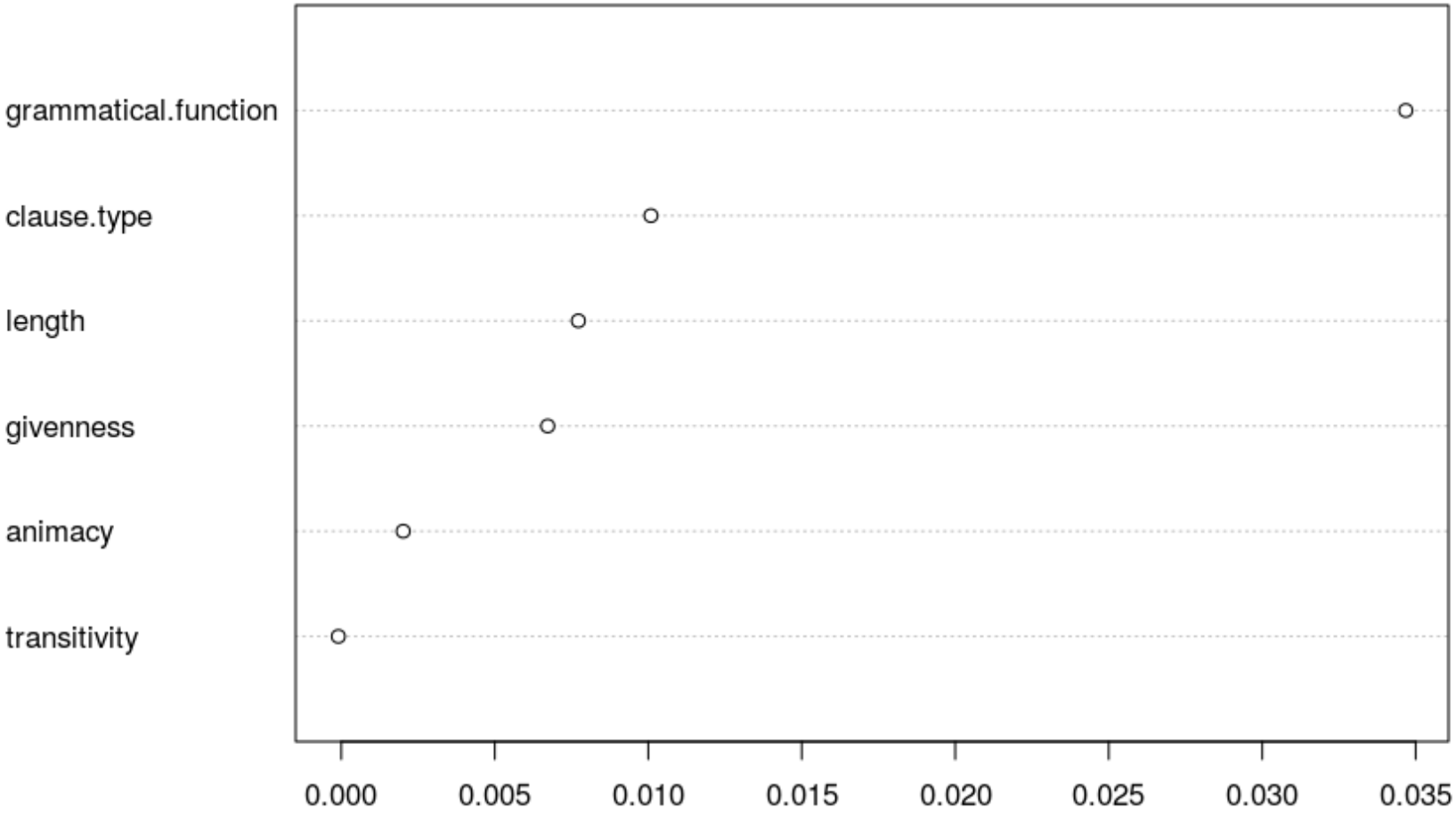
- Conditional inference trees and random forests:
  - explore interactions between predictors
  - robustness to correlated predictors
- Details:
  - `ctree`, `cforest` from `party`
  - forests: 300 trees, mtry = 3
  - confusion matrix and accuracy based on OOB predictions

Grammatical function: conditional inference tree



position ~ animacy + clause.type + givenness + grammatical function + length + transitivity

Grammatical function: random forest

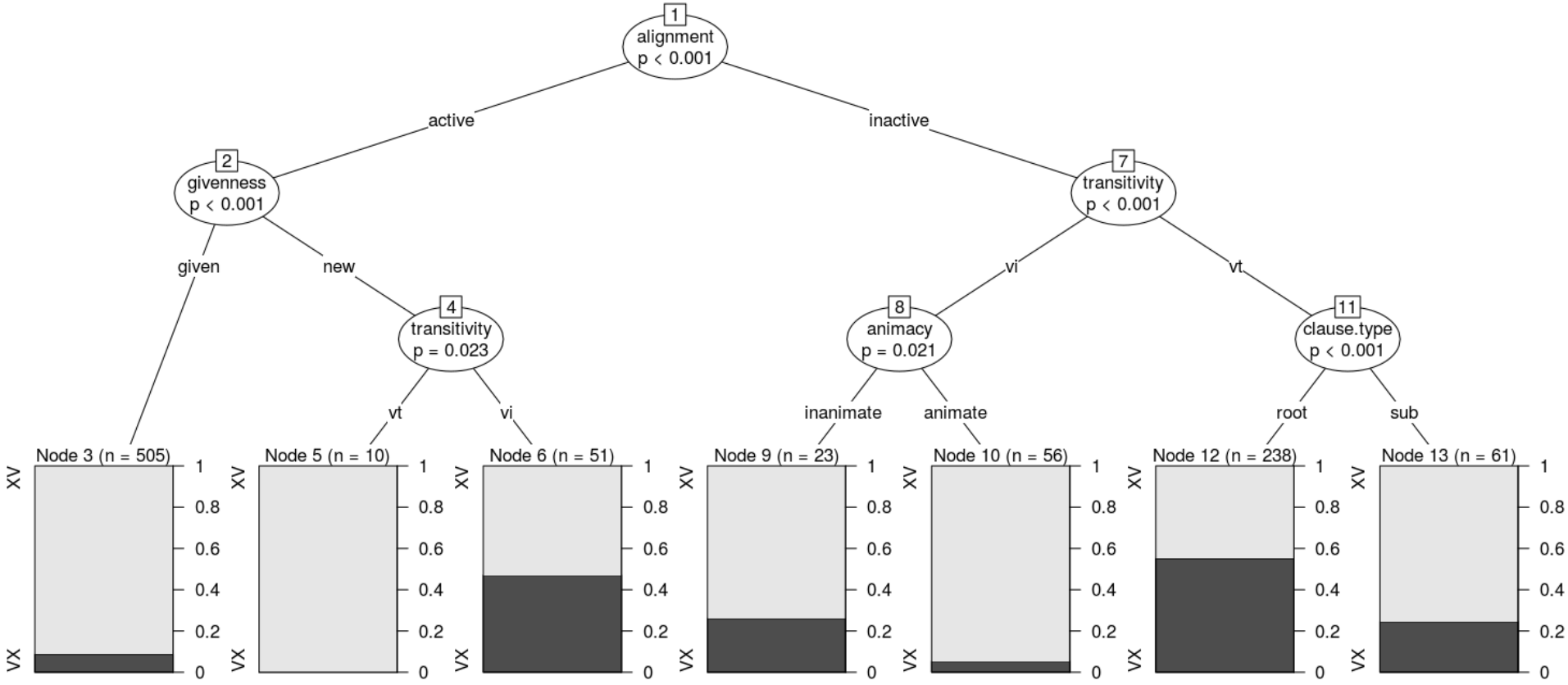


Accuracy: 78.4%

Baseline: 76.3%

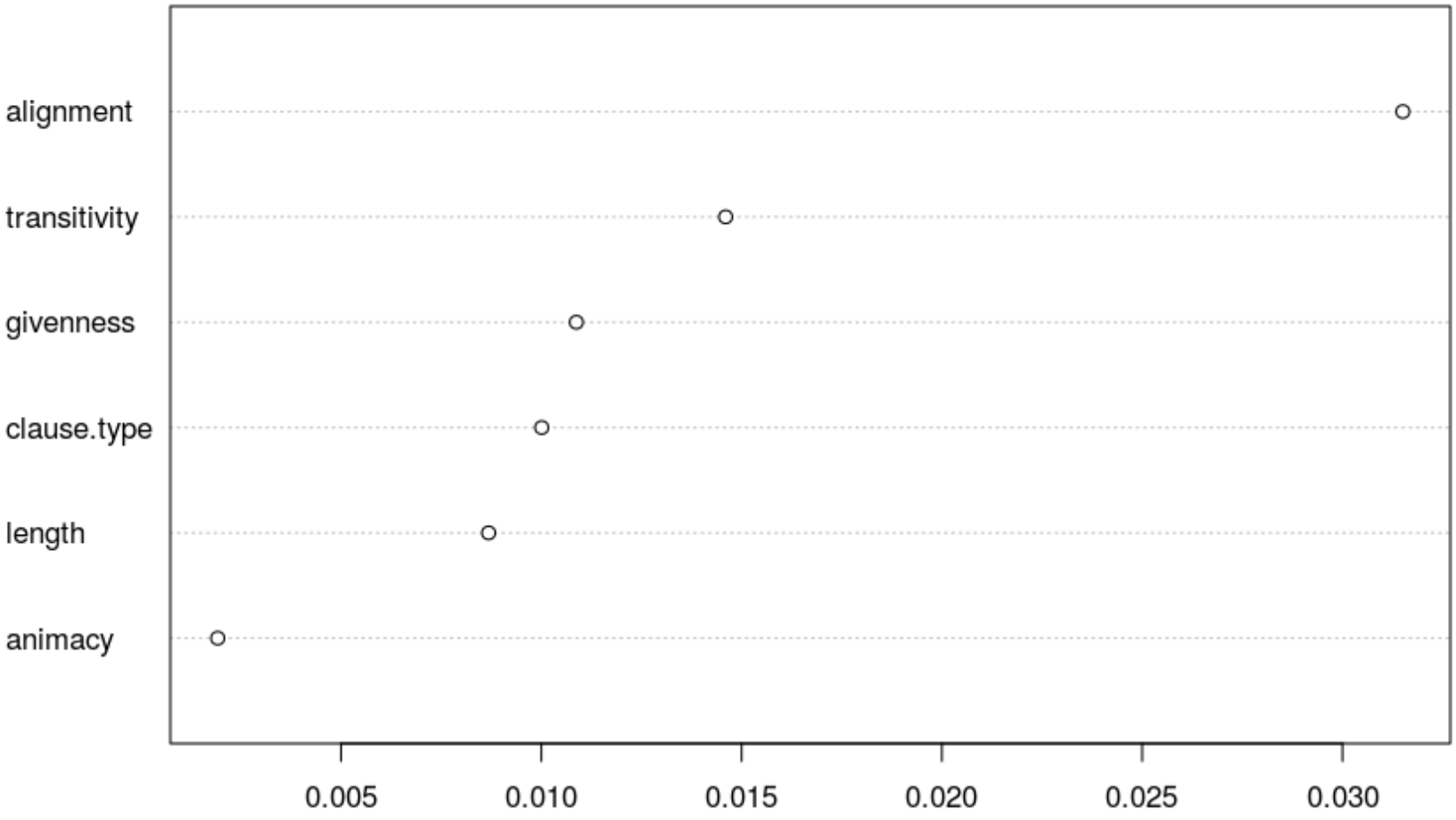
	XV	VX
XV	655	66
VX	138	85

(In)active alignment: conditional inference tree



position ~ alignment + animacy + clause.type + givenness + length + transitivity

(In)active alignment: random forest



Accuracy: 77.9%

Baseline: 76.3%

	XV	VX
XV	654	67
VX	142	81

Zooming in on intransitive verbs

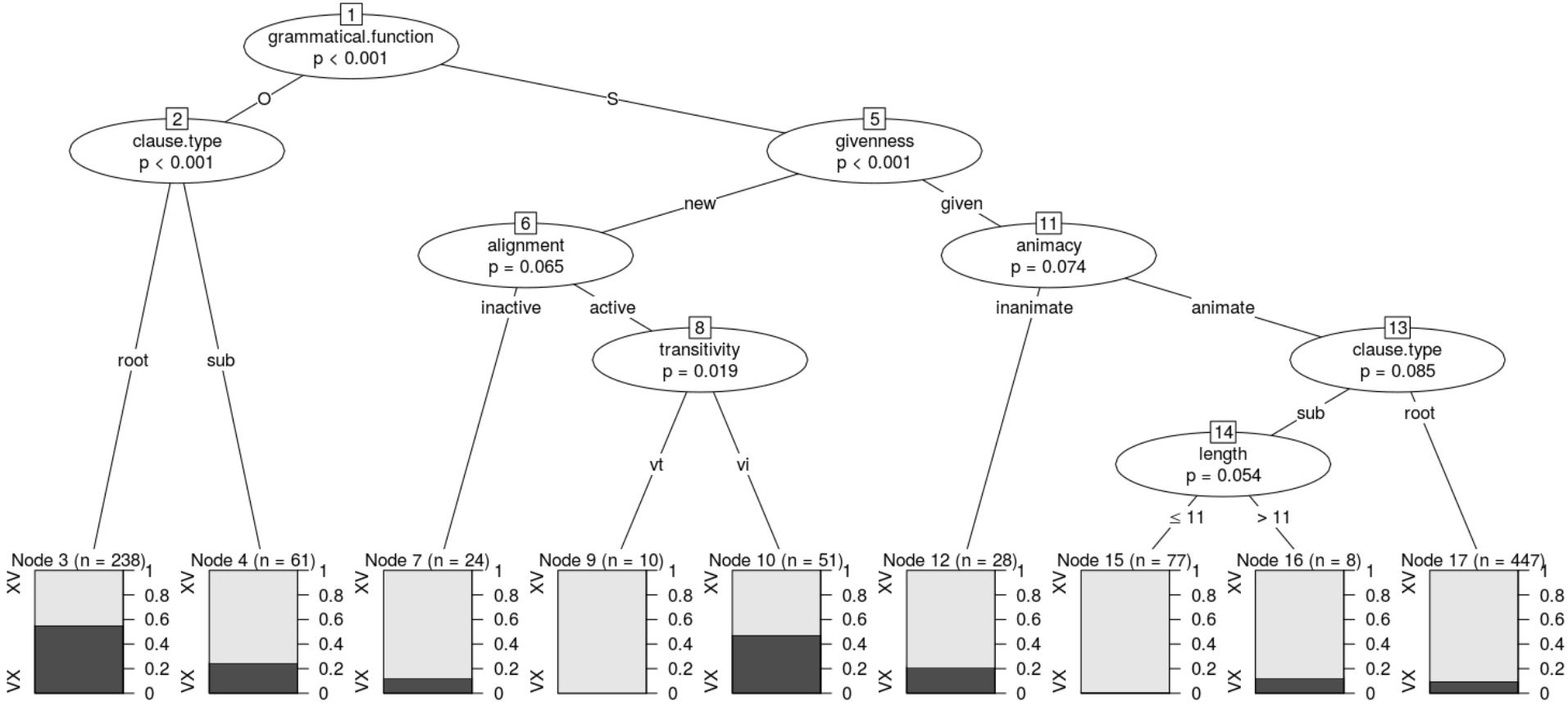
- New active intransitive Ss more likely preverbal than other Ss
- 82% verbs of location, movement, perception and existence:

Lemma	Translation	freq
ĩ	be present	8
iko	exist	18
japukai	shout	2
jekuaa	appear	1
nhe'ẽ	speak	3
nhendu	be heard	5

Lemma	Translation	freq
o	go	3
pẽ	break	1
u	come	4
vaẽ	arrive	3
vy'a	rejoice	3

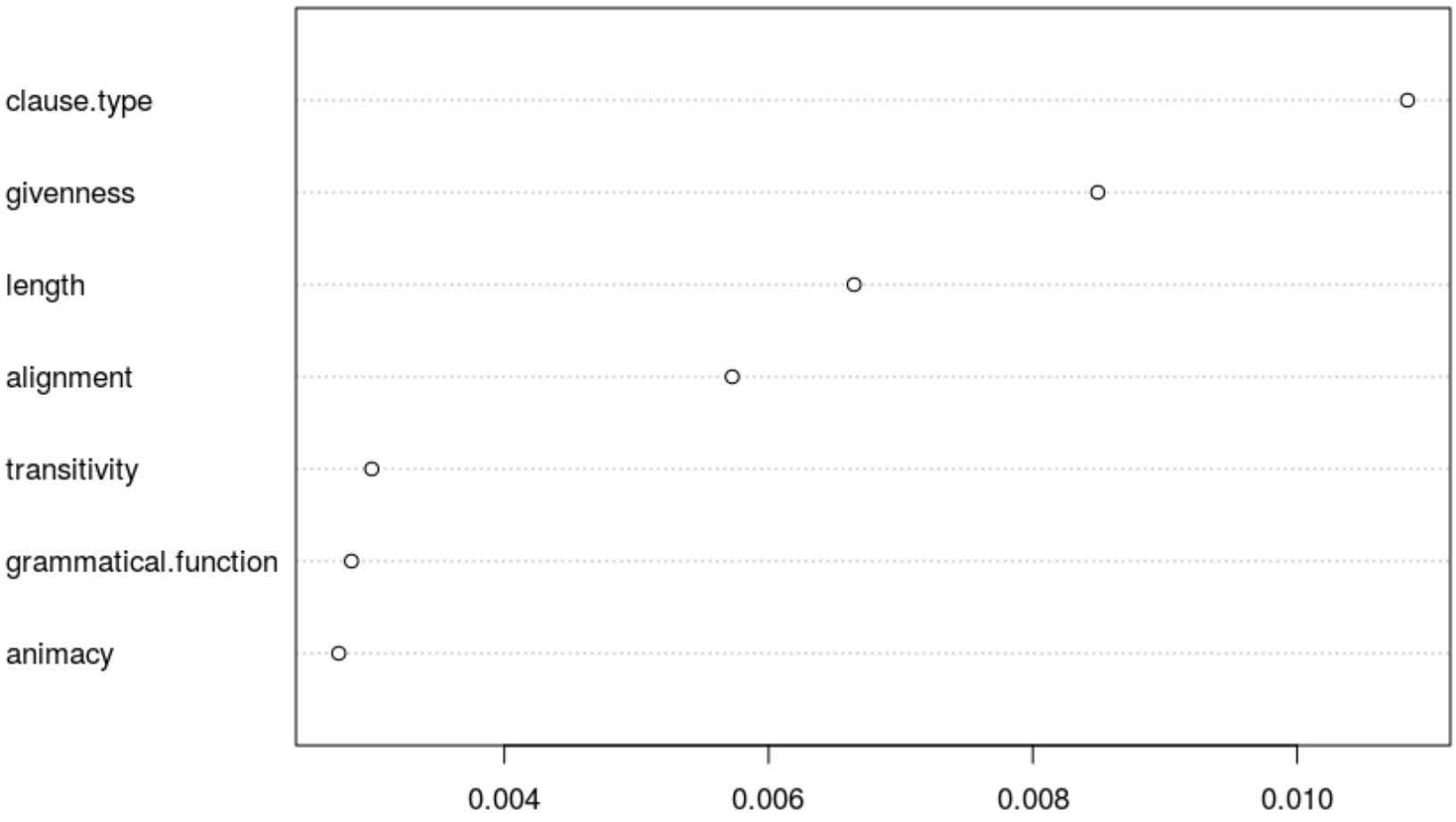
- Source arguments coded as actors (Velazquez Castillo 2002)
- Hypothesis: presentative/directive inversions

Complete model



position ~ alignment + animacy + clause.type + givenness + grammatical.function + length

Complete model



Accuracy: 78.6%

Baseline: 76.3%

	XV	VX
XV	651	70
VX	132	91

Discussion

S/O description of argument position in Mbyá

- Dominantly SVO
- Dominantly SV (88.1%)
- No dominant OV/VO order (51.2% preverbal)
- Subordinate O more likely preverbal than root O (75.4% vs 45%)
- Given arguments more likely preverbal than new ones (81.8% vs 57.7%)

A/P description of argument position in Mbyá

- Dominantly AVP
- Dominantly AV (88%)
- **Dominantly PV** (59%)
- Subordinate P more likely preverbal than root P (82% vs 51.9%)
- **Transitive P more likely postverbal that intransitive P** (48.4% vs 11.4%)
- Given arguments more likely preverbal than new ones (81.8% vs 57.7%)

Taking stock

- For word order typologies, either description appear to be reasonable
- For multifactorial models, no reason not to include both factors in models where collinearity is not an issue
- Grammatical function is more strongly associated with argument order than alignment
- Interesting interaction between alignment, givenness and transitivity

Mbyá word order in perspective

- Tonhauser & Colijn (2010), word order in Paraguayan Guaraní
  - 2,800 words corpus, only matrix clauses
  - 55% preverbal subjects, 95% postverbal objects
- AILLA corpus, matrix clauses:
  - 86.8% preverbal subjects, 55% postverbal objects
- OV $\rightarrow$ VO evolution in Tupí-Guaraní (Dietrich 2009)
  - subordinate clauses more conservative (Bybee 2002)
  - Paraguayan Guaraní more in contact with Spanish

Thank You