There are patterns in child-directed speech that make first language learning possible

Sabine Stoll

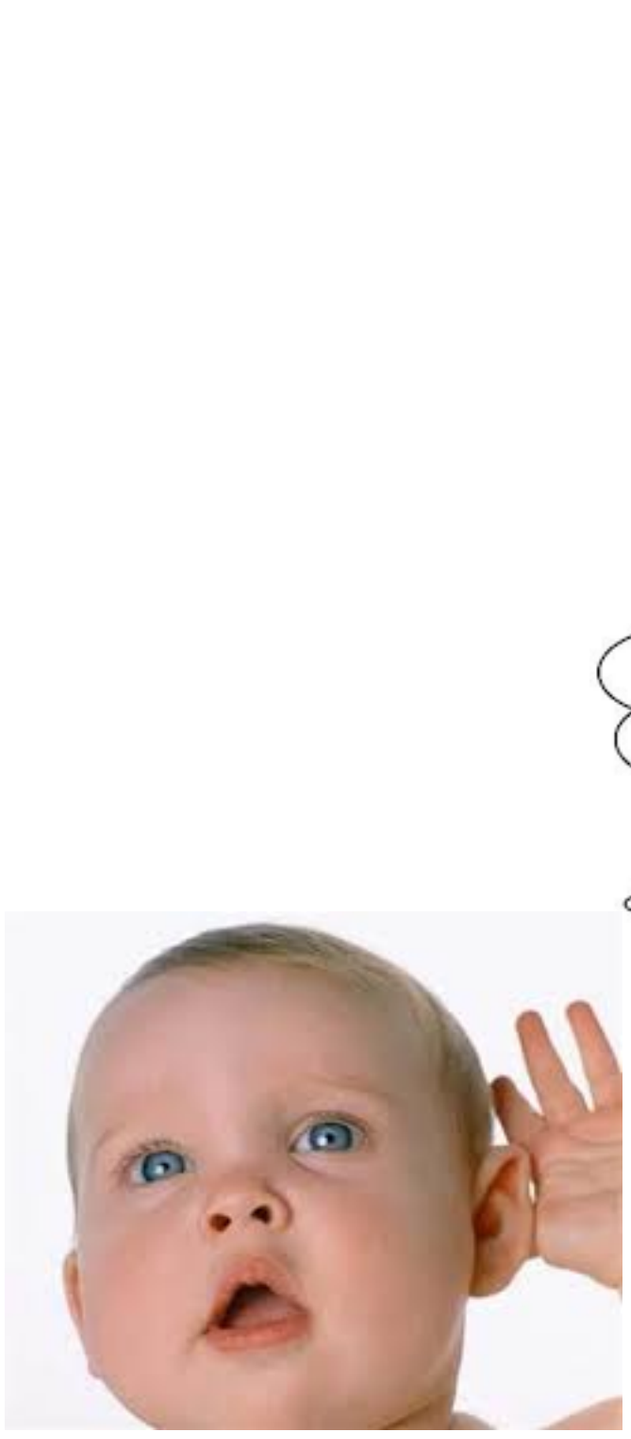
University of Zurich
Department of Comparative Linguistics
Psycholinguistics Laboratory
@sabine_stoll
&
Center for the Interdisciplinary Study of Language Evolution (ISLE)
@ISLE_uzh

University of Zurich[UZH]

FNSNF
Swiss National Science Foundation

erc

How do children learn language?



My sweet baby

Recursion

Recursion

Rule-based learning

house	**houses**	**child**	**children**
kid	**kids**	**mouse**	**mice**
dog	**dogs**	**foot**	**feet**

Rule-based learning

German

Table 1: The Eight German Plural Classes

affix	-s	-(e)n	-e	-er	zero
- Umlaut	*Auto-s*	*Bahn-en, Fahne-n*	*Fisch-e*	*Kind-er*	*Lehrer-0*
+ Umlaut	--	--	*Bä̲nk-e*	*Rä̲d-er*	*Kä̲sten-0*

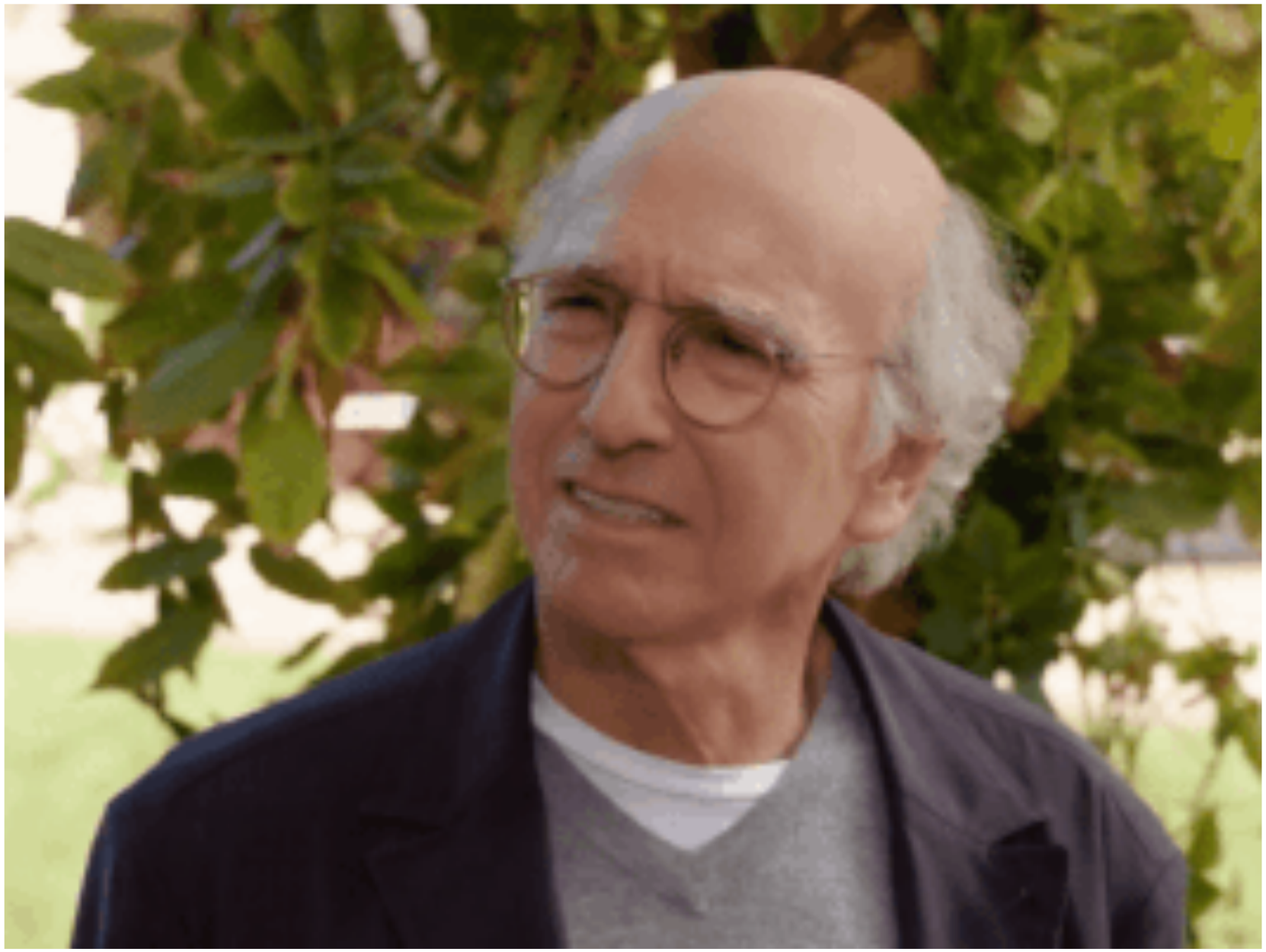
Behrens 2002

Dinka

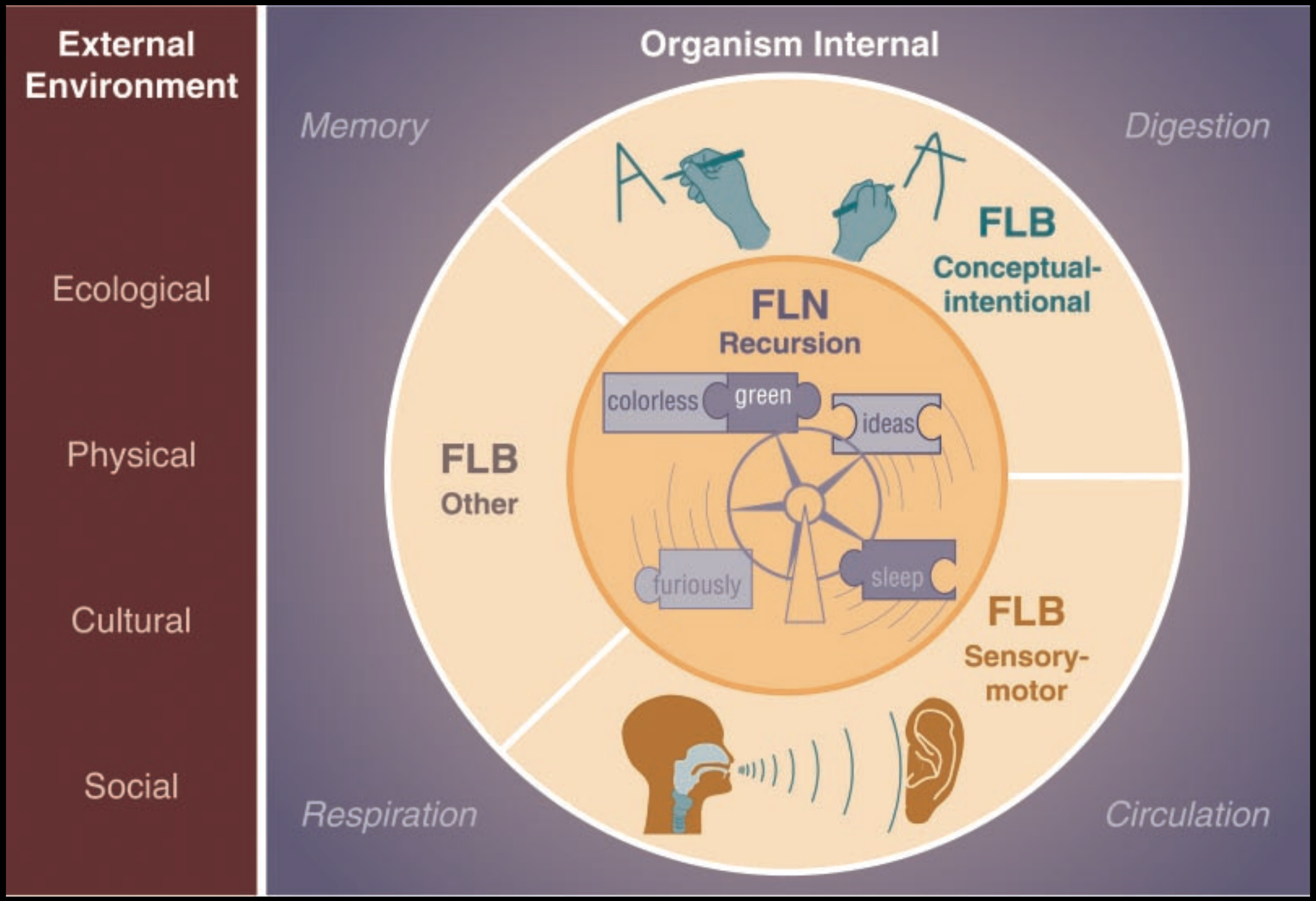
SINGULAR	PLURAL	GLOSS	
cǐin	cìn	'hand'	(marked singular)
jɛ̤́ɛc	jɛ̤̀eec	'belly'	(marked plural)
tṳ̀uk	tṳ́uk	'fruit of palm'	(both singular and plural marked)

Beyond structure

Beyond structure

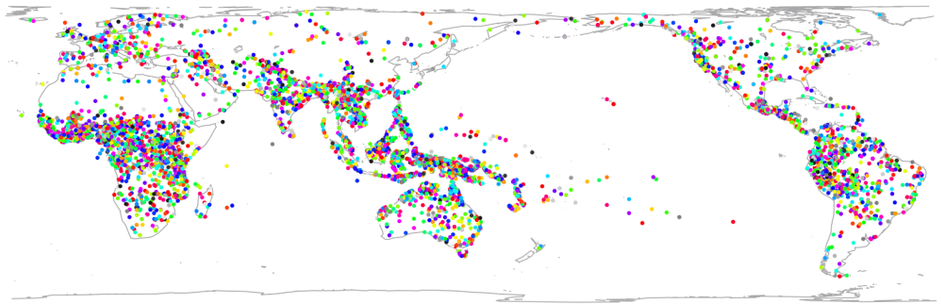


Beyond structure



Hauser, Chomsky & Fitch 2002

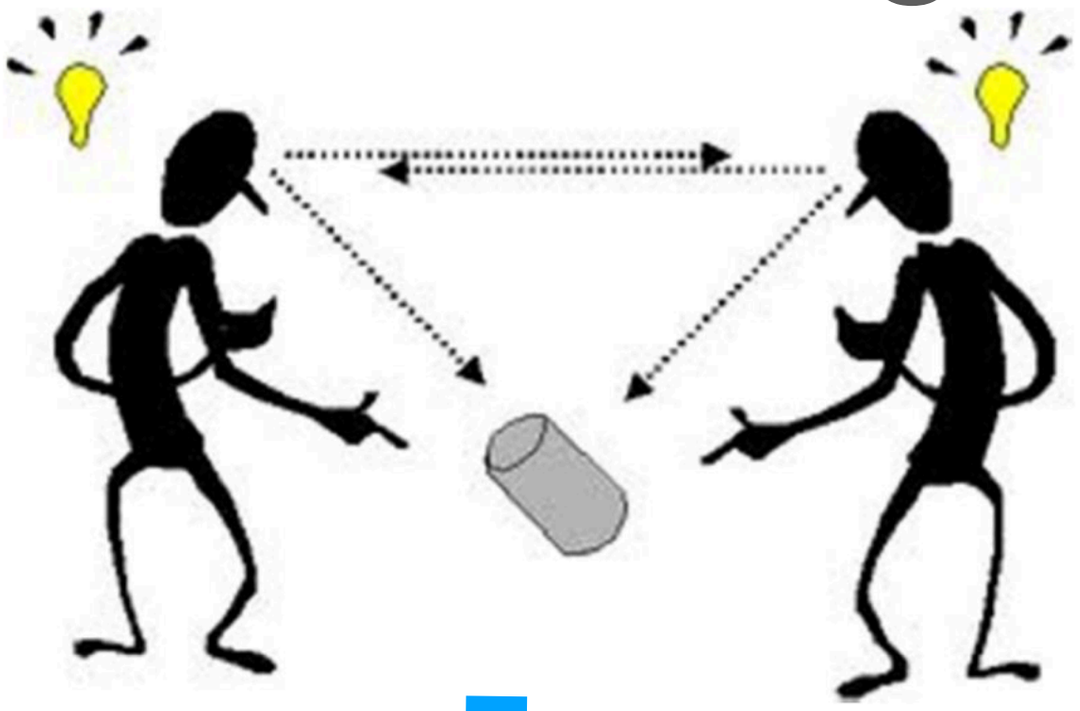
Learning mechanisms

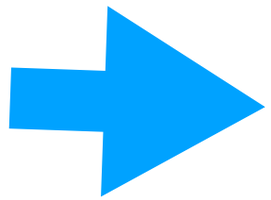


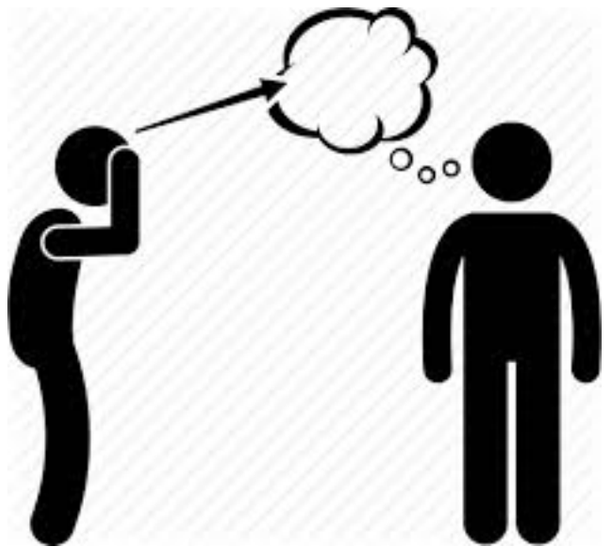
Socio-cognitive abilities

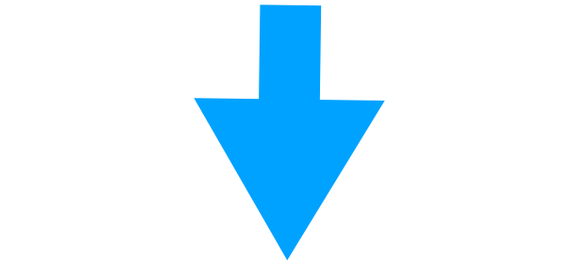
Experiments

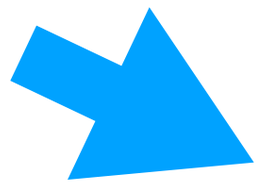
Learning mechanisms

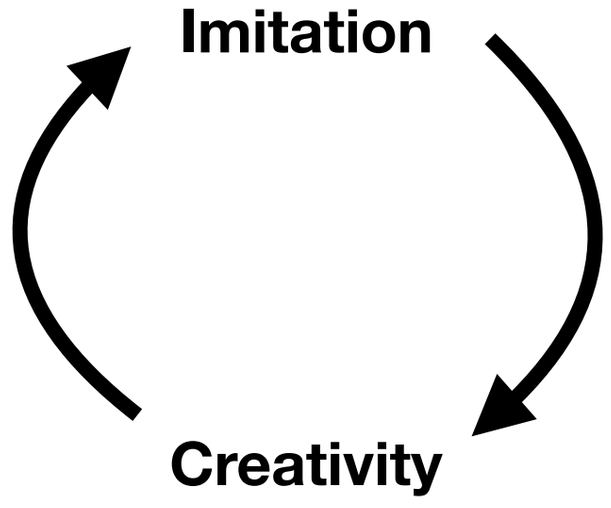




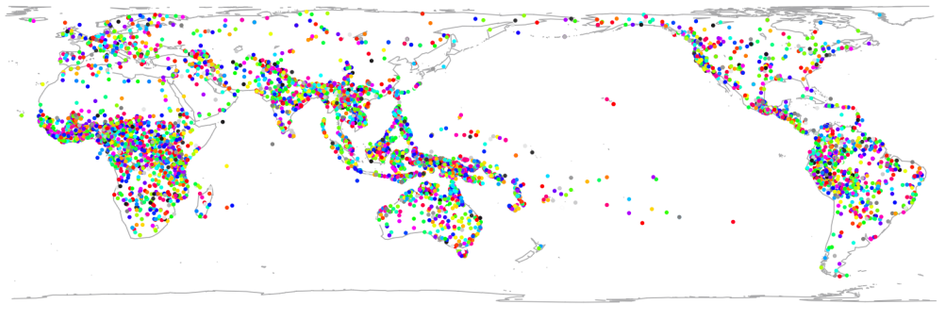








Learning mechanisms



Socio-cognitive abilities

Experiments

Pattern-detection abilities

Experiments

How is language learning possible?

- **Study 1:** An extreme case of language learning: A polysynthetic verb system.
- **Study 2**: Variation sets: How to learn about constructions and meaning.
- **Study 3**: Structural frames: how to detect nouns and verbs
- **Study 4**: Temporal cues: how to detect nouns and verbs

Balthasar Bickel

Damian Blasi

Jekaterina Mazara

Study 1: An extreme case of language learning: thousands of verb forms in Chintang

Verbal morphology

Chintang

English

I walk
you walk
he walks
we walk
you walk
they walk
walked

	1s	1di	1pi	1de	1pe	2s	2d	2p	3s	3ns	intransitive
1s						*tupnaʔā* *tupnaʔāniŋ* *tupnehẽ* *matupyoknehẽ*	*tupnaʔãce* *tupnaʔãceniŋ* *tupnace* *matupyoknace*	*tupnaʔāni* *tupnaʔāniniŋ* *tupnanihẽ* *matupyoknanihẽ*	*tubukuŋ* *tubukuŋniŋ* *tubuhẽ* *matupyoktuhẽ*	*tubukuŋcuŋ* *tubukuŋcuŋniŋ* *tubuŋcihẽ* *matupyoktuŋcihẽ*	*tupmaʔā* *tupmaʔāniŋ* *tubehẽ* *matupyoktehẽ*
1di									*tupcoko* *tupcokoniŋ* *tubace* *matupyoktace*	*tubumcum* *tubumcumnim* *tubumcumhẽ* *matupyoktumcumhẽ*	*tupceke* *tupcekeniŋ* *tubace* *matupyoktace*
1pi									*tubukum* *tubukumnim* *tubumhẽ* *matupyoktumhẽ*		*tubiki* *tubikiniŋ* *tubihẽ* *matupyoktihẽ*
1de							*tupnaʔãncĩyã* *tupnaʔãncĩyãniŋ* *tupnancĩyehẽ* *matupyoknancĩyehẽ*		*tupcokoŋa* *tupcokoŋaniŋ* *tubacehẽ* *matupyoktacehẽ*	*tubumcumma* *tubumcummaniŋ* *tubumcummehẽ* *matupyoktumcummehẽ*	*tupcekeŋa* *tupcekeŋaniŋ* *tubacehẽ* *matupyoktacehẽ*
1pe									*tubukumma* *tubukummaniŋ* *tubummehẽ* *matupyoktummehẽ*		*tubikiŋa* *tubikiŋaniŋ* *tubiehẽ* *matupyoktiehẽ*
2s	*atupmaʔā* *atupmaʔāniŋ* *atubehẽ* *{a-ma}tupyoktehẽ*			*{a-ma}tupceke* *{a-ma}tupcekeniŋ* *{a-ma}tubace* *{a-ma-ma}tupyoktace*	*{a-ma}tupno* *{a-ma}tupnikniŋ* *{a-ma}tube* *{a-ma-ma}tupyokte*				*atubokoniŋ* *atube* *amatupyokte*	*atubukuceniŋ* *atubuce* *{a-ma}tupyoktuce*	*atupno* *atupniknŋ* *atube* *{a-ma}tupyokte*
2d	*atupmaʔanciŋ* *atupmaʔanciŋniŋ* *atubaŋcihẽ* *{a-ma}tupyoktaŋcihẽ*								*atupcoko* *atupcokoniŋ* *atubace* *amatupyoktace*	*atubumcum* *atubumcumnim* *atubumcumhẽ* *{a-ma}tupyoktumcumhẽ*	*atupceke* *atupcekeniŋ* *atubace* *{a-ma}tupyoktace*
2p	*atupmaʔaniŋ* *atupmaʔaniniŋ* *atubaŋnihẽ* *{a-ma}tupyoktaŋnihẽ*								*atubukum* *atubukumnim* *atubumhẽ* *amatupyoktumhẽ*		*atubiki* *atubikiniŋ* *atubihẽ* *{a-ma}tupyoktihẽ*
3s	*utupmaʔā* *utupmaʔāniŋ* *utubehẽ* *{u-ma}tupyoktehẽ*	*maitupceke* *maitupcekeniŋ* *maitubace* *{mai-ma}tupyoktace*	*maitupno* *maitupnikniŋ* *maitube* *{mai-ma}tupyokte*	*matupceke* *matupcekeniŋ* *matubace* *{ma-ma}tupyoktace*	*matupno* *matupnikniŋ* *matube* *{ma-ma}tupyokte*	*natupno* *natupnikniŋ* *natube* *{na-ma}tupyokte*	*natupceke* *natupcekeniŋ* *natubace* *{na-ma}tupyoktace*	*natubiki* *natubikiniŋ* *natubihẽ* *{na-ma}tupyoktihẽ*	*tuboko* *tubokoniŋ* *tube* *matupyokte*	*tubukuce* *tubukuceniŋ* *tubuce* *matupyoktuce*	*tupno* *tupnikniŋ* *tube* *matupyokte*
3d	*utupmaʔanciŋ* *utupmaʔanciŋniŋ* *utubaŋcihẽ* *{u-ma}tupyoktaŋcihẽ*								*utupcoko* *utupcokoniŋ* *utubace* *{u-ma}tupyoktace*	*utubukuce* *utubukuceniŋ* *utubuce* *{u-ma}tupyoktuce*	*utupceke* *utupcekeniŋ* *utubace* *{u-ma}tupyoktace*
3p	*utupmaʔaniŋ* *utupmaʔaniniŋ* *utubaŋnihẽ* *{u-ma}tupyoktaŋnihẽ*								*utuboko* *utubokoniŋ* *utube* *{u-ma}tupyokte*		*utupno* *utupnikniŋ* *utube* *{u-ma}tupyokte*

jo-go-yaŋ *na-khutt-i-**ca**-i-**hatt**-i-**bir**-i.*

whatever-NMLZ-ADD 3[s]>2-steal-2pO-V2:eat-2pO-V2:move.away.TR-2pO-V2:do.for-[SBJV.]2pO

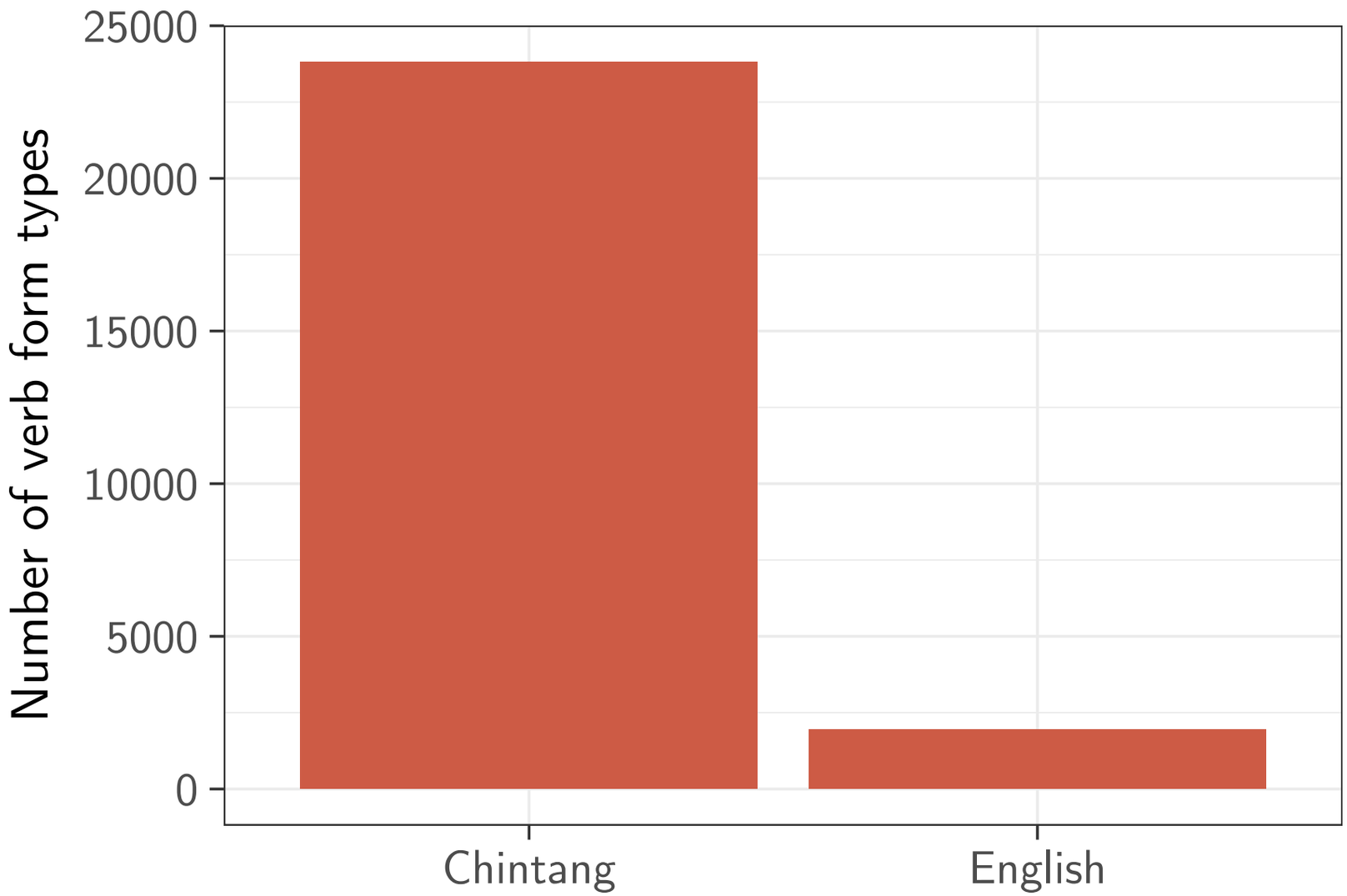
‘It (a cat) may steal everything from you and eat it all up!’ [story.cat.204]

Verbal structure

	Affixes	Unique synthetic forms
English	3	3
Chintang	148	4745

How can such an extreme system be learned?

Verb form types in the input

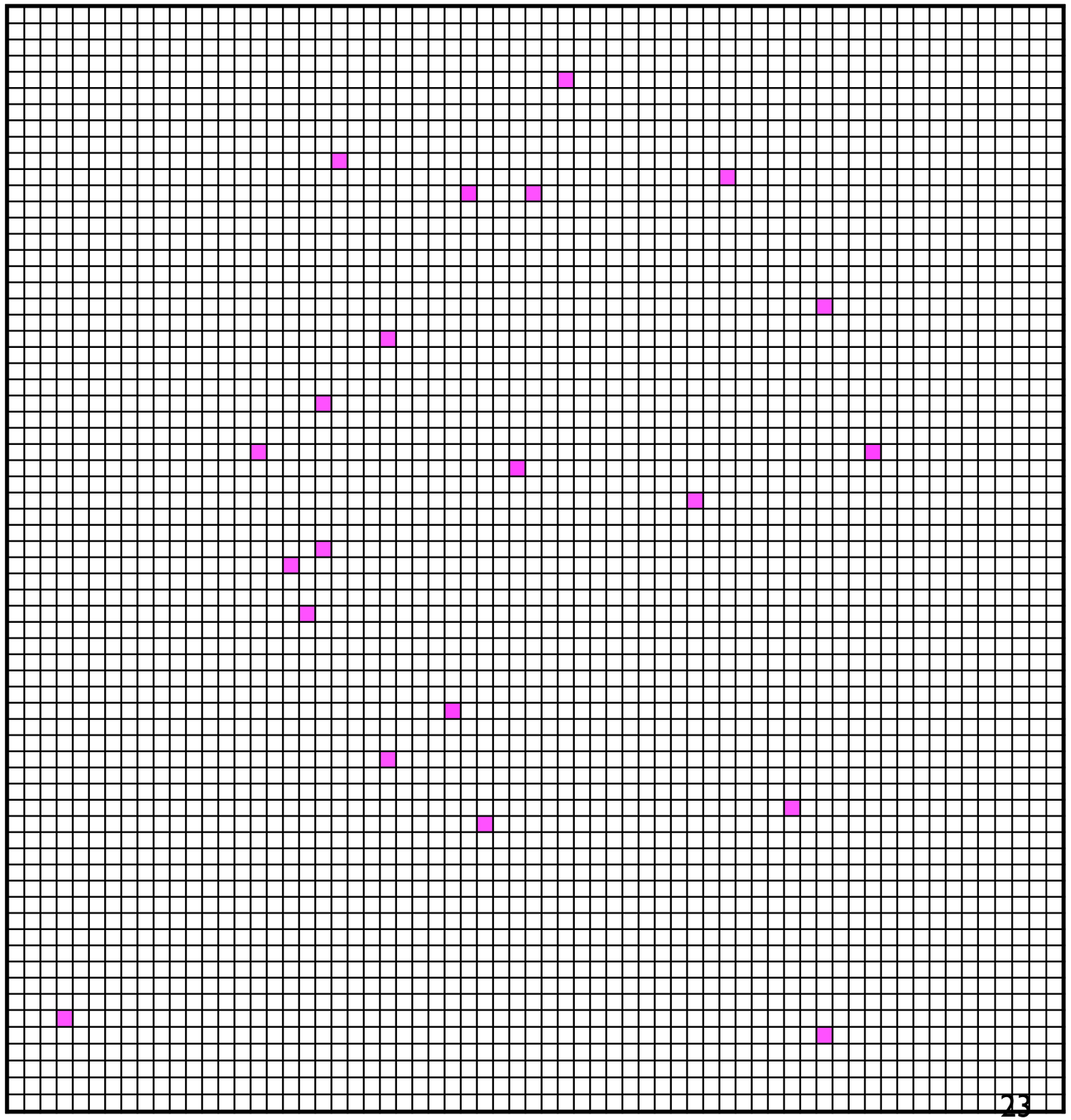


English

Chintang

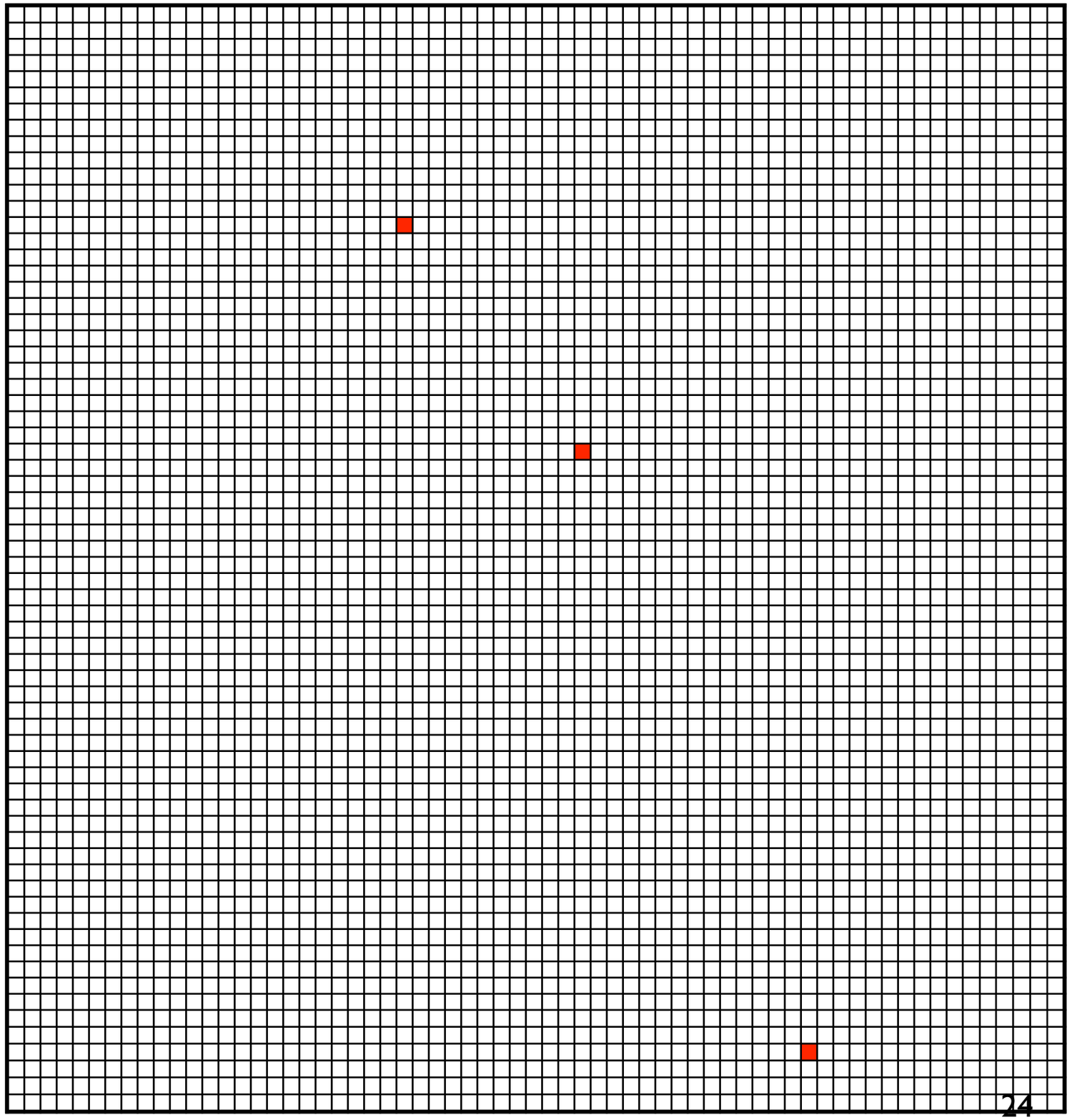
English

Chintang



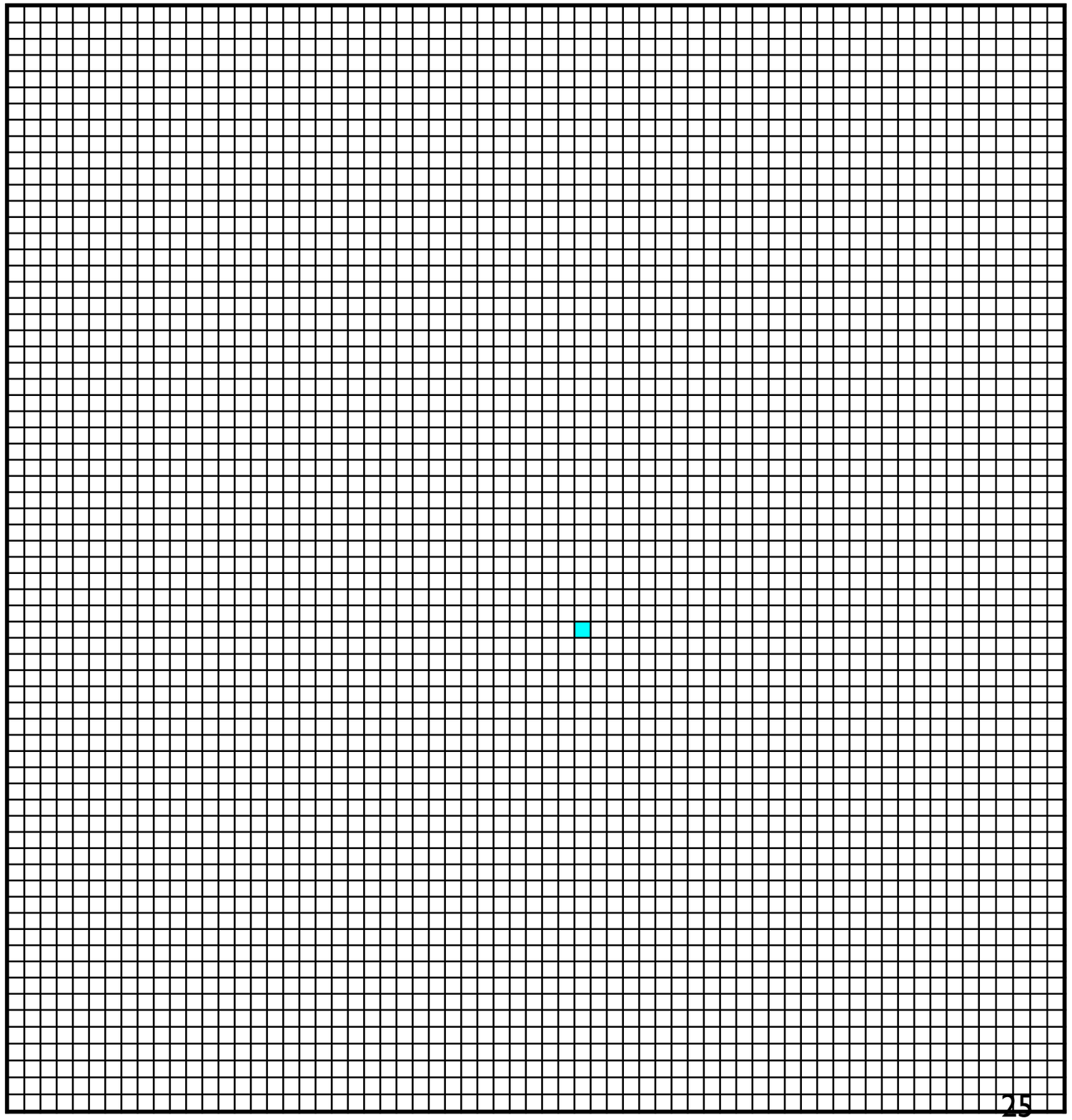
English

Chintang

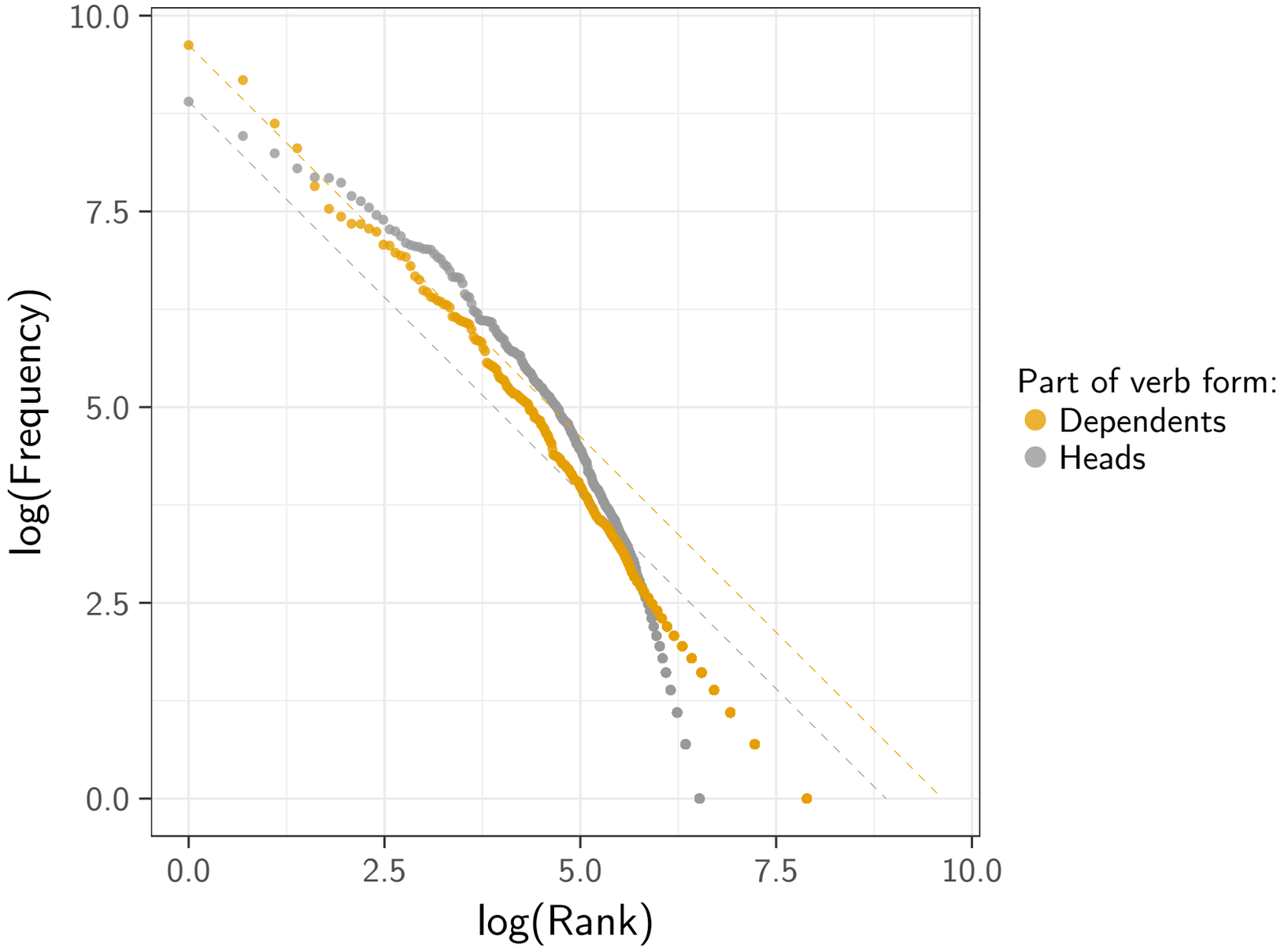


English

Chintang

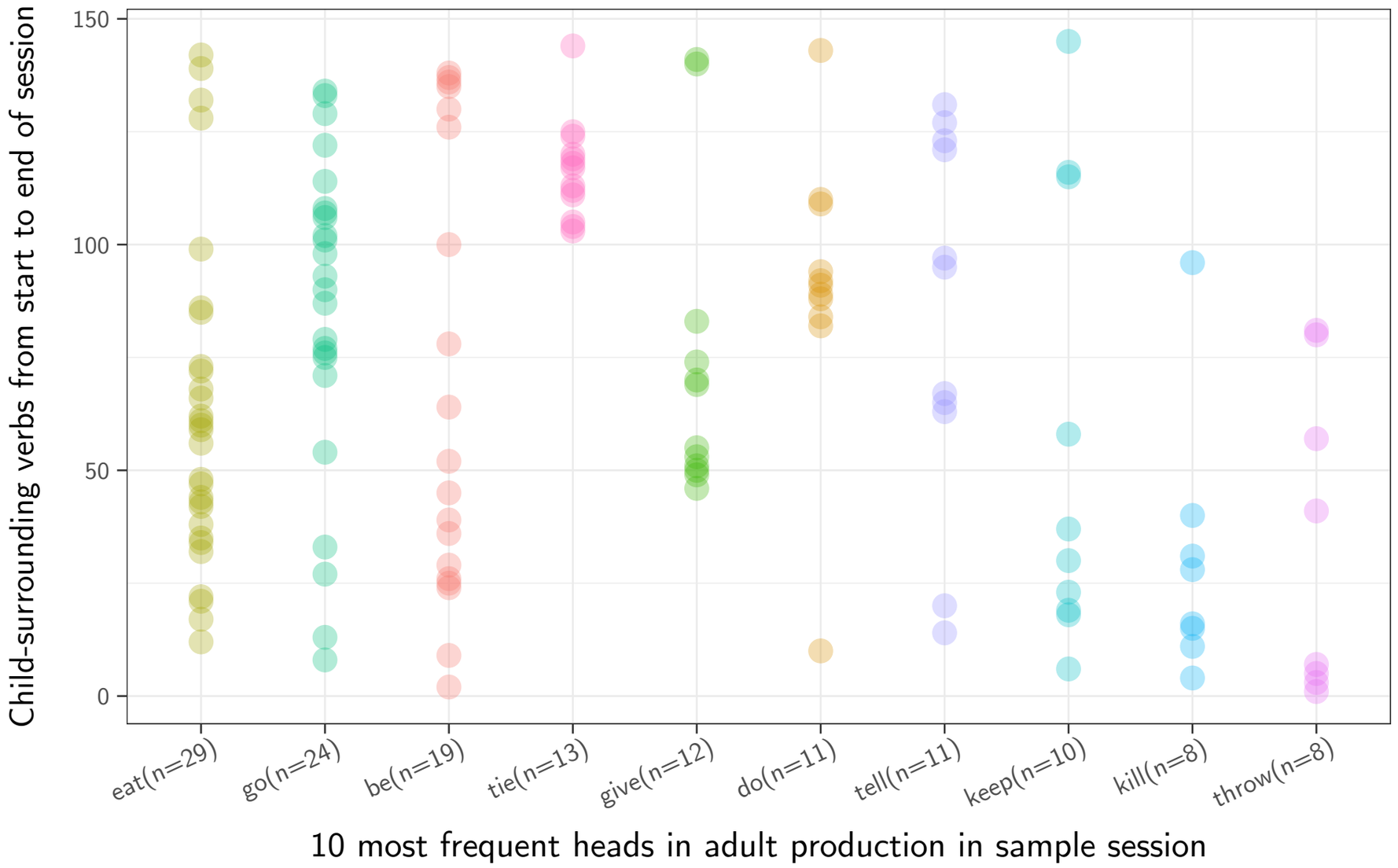


Frequency distributions stems vs. affixes: Chintang (input)



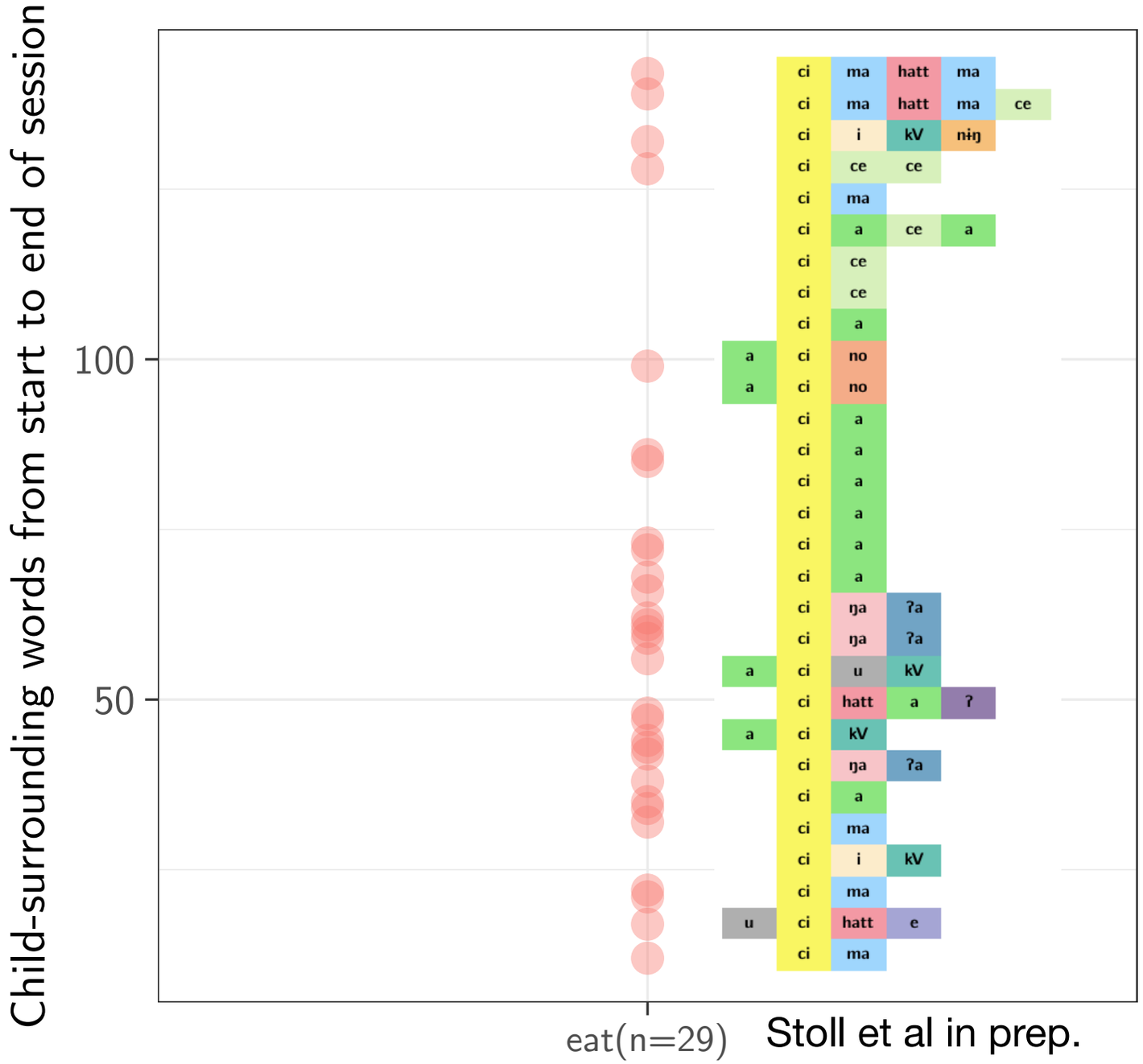
Stoll et al in prep.

Most frequent stems in one recording: Chintang (input)

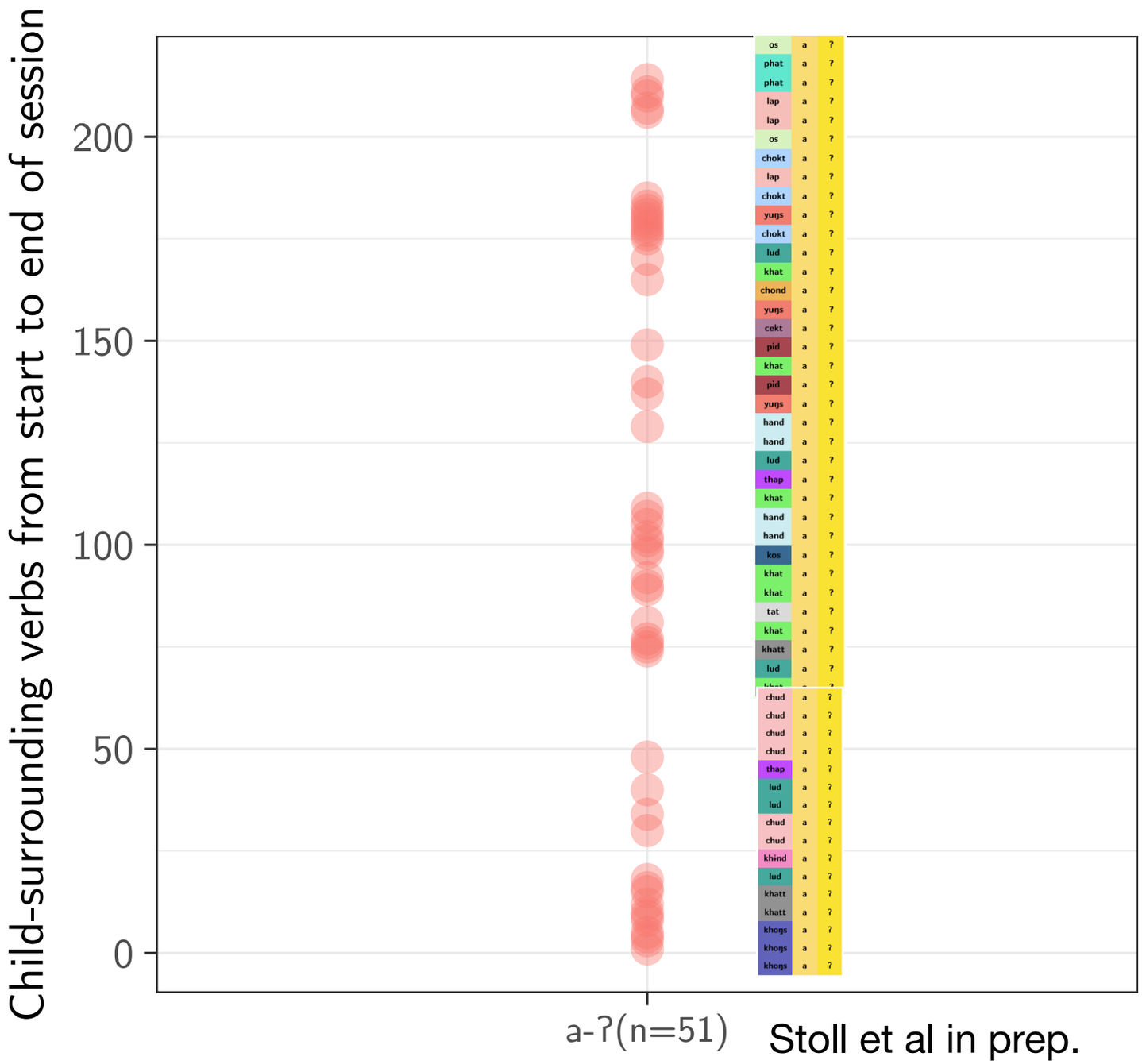


Stoll et al in prep.

Forms with the same stem in one recording session: Chintang (input)

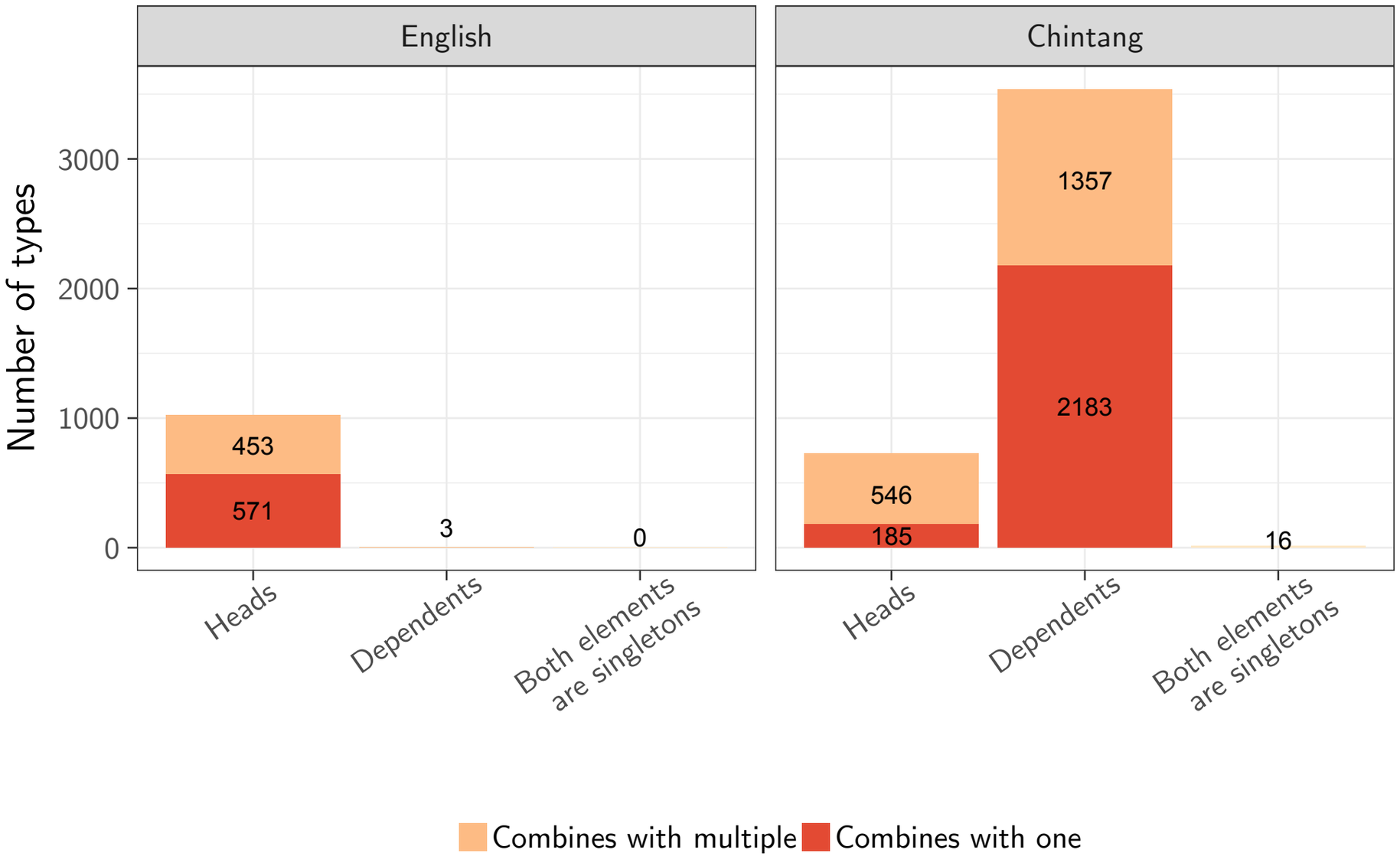


Forms with same affixes in one recording session: Chintang input



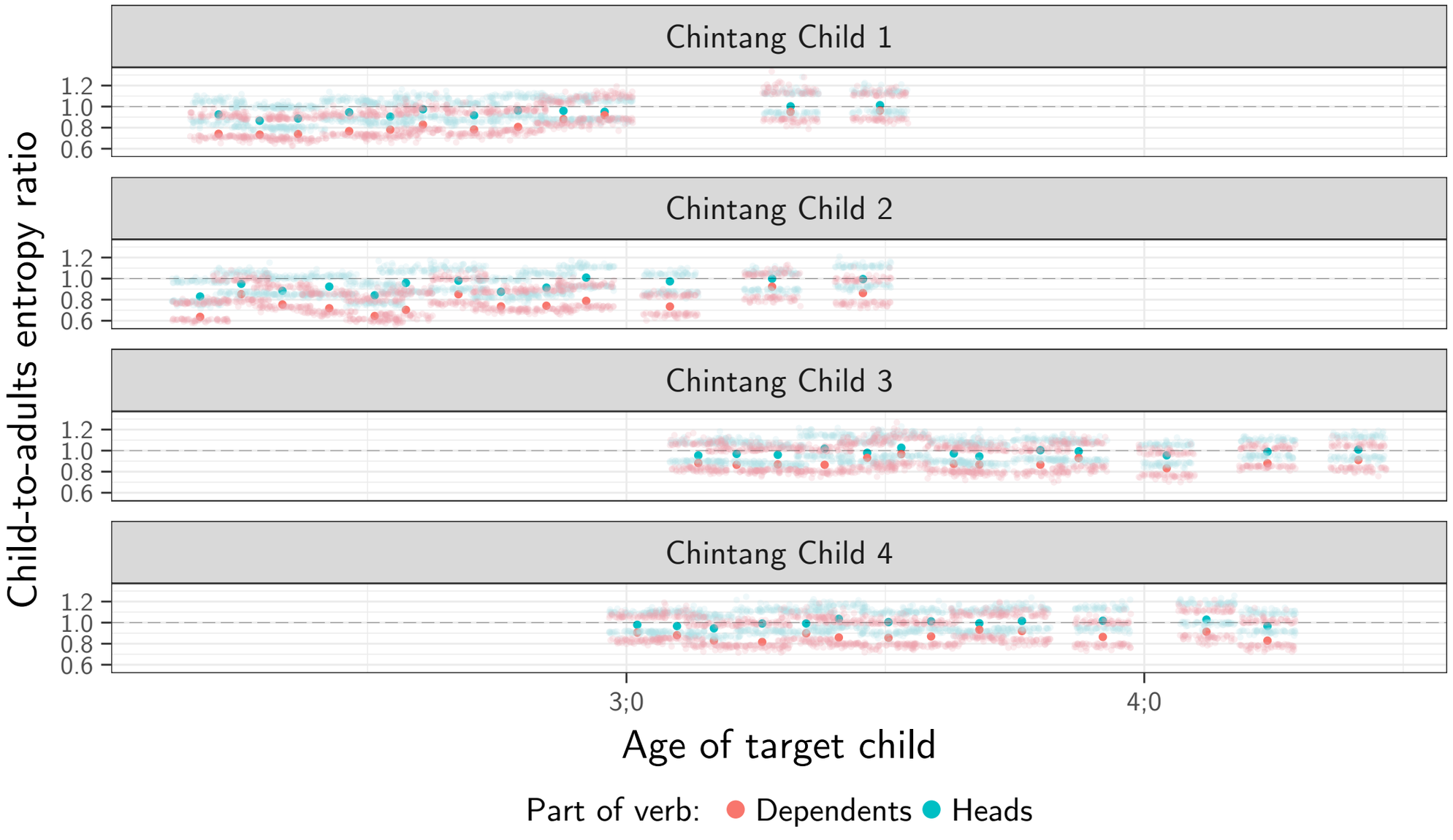
Stoll et al in prep.

Stem and affix combinations: Chintang (input)



Stoll et al in prep.

Stems vs. Affixes over time in children

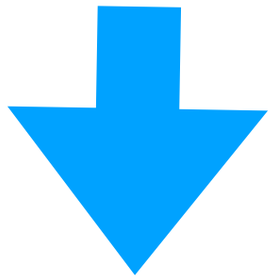


Stoll et al in prep.

Conclusion
Chintang verbal morphology

The system can be learned via distributional learning, generalisation and abstraction

How do children learn language?



How come children can learn **any** language in a few years time?

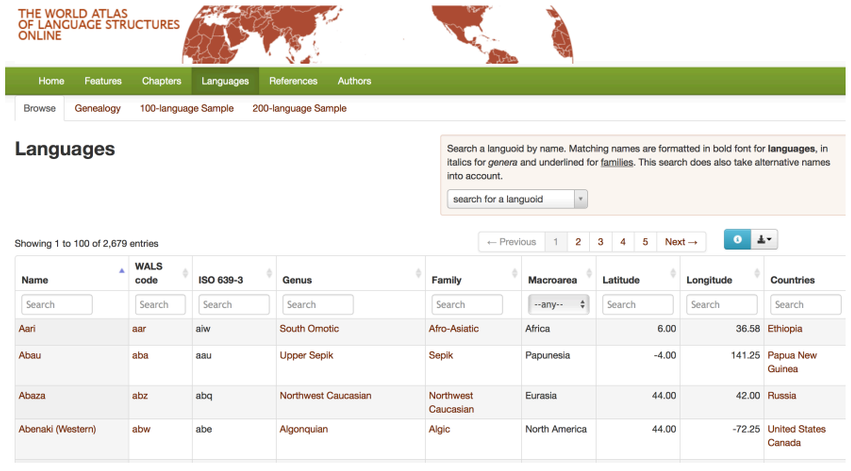
Challenge: diversity

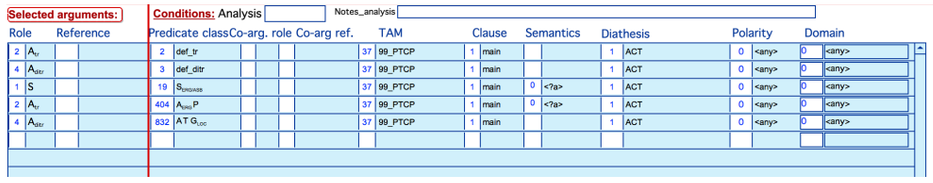
Data

Grammars

WALS

AUTOTYP

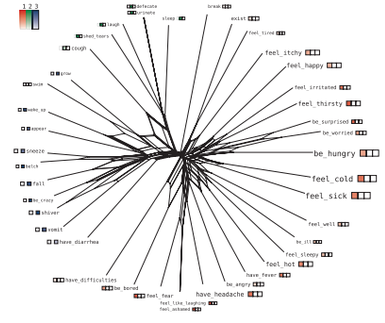





```
group_by(Language, ReferenceType, TAM, Polarity, Diathesis) %>%
arrange(Role) %>%
do({
  # split the roles by GRs
  roles <- lapply(split(.$Role, .$GR, drop=T), as.character)

  # convert them to alighment, sorting appropriately
  eqset <- sapply(roles, paste, collapse='=')
  eqset <- factor(eqset, levels=c('S', 'S=Atr=P', 'S=A', 'S=P', 'Atr', 'Atr=P', 'P'))
  eqset <- as.character(sort(eqset))
  alignment <- paste(eqset, collapse='≠')
  data.frame(Alignment = alignment, stringsAsFactors=F)
})
```

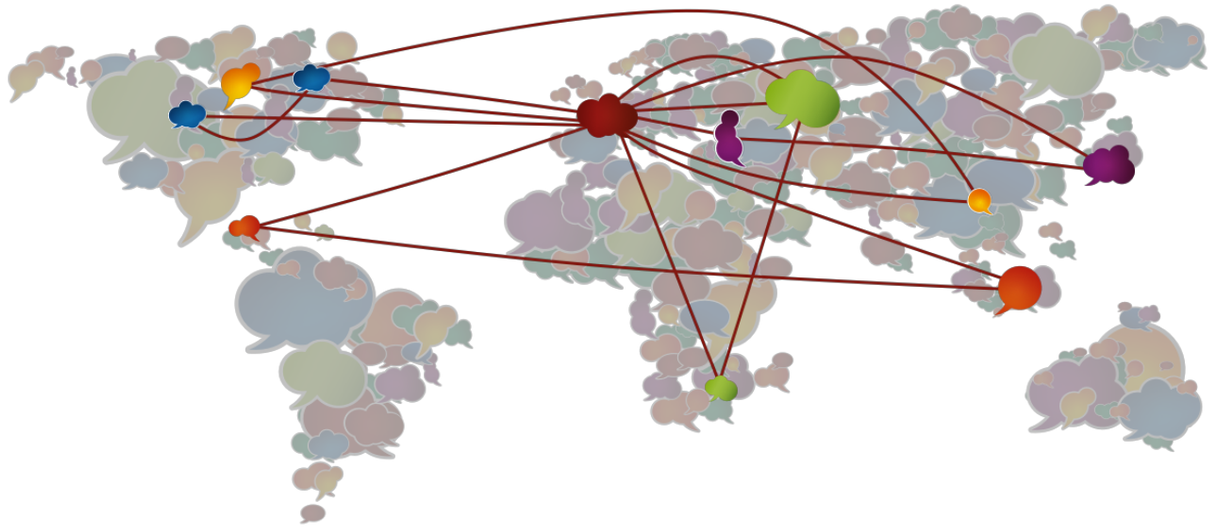
```
build_mat <- function(dta)
{
items <- unique(unlist(dta))
markers  <- names(dta)

mat <- matrix(0, nrow=length(items), ncol=length(markers))
colnames(mat) <- markers
rownames(mat) <- items

for(m in markers) for(i in items) if(i %in% dta[[m]]) mat[i, m] <- 1

mat
}
# S class
rolemat5 <- build_mat(semantics.per.role['S'])
```

Language	ReferenceType	TAM	Polarity	Diathesis	Alighment
Abkhaz	#ANY#	<NA>	<NA>	ACT	S=Atr=P
Acehnese	#ANY#	<NA>	<NA>	ACT	S=Atr=P
Ainu	#ANY#	<NA>	<NA>	ACT	S=Atr=P
Alamblak	#ANY#	<NA>	<NA>	ACT	S=Atr=P
Akkadian	#ANY#	<NA>	<NA>	ACT	S=Atr≠P
Albanian	2plPro	<NA>	<NA>	ACT	S=Atr=P
Albanian	N-def	<NA>	<NA>	ACT	S=Atr≠P
Albanian	3sgPro	<NA>	<NA>	ACT	S=Atr≠P
Albanian	1sgPro	<NA>	<NA>	ACT	S=Atr≠P
Albanian	2sgPro	<NA>	<NA>	ACT	S=Atr≠P
Albanian	N-indef	<NA>	<NA>	ACT	S=Atr=P
Albanian	1plPro	<NA>	<NA>	ACT	S=Atr=P
Albanian	3plPro	<NA>	<NA>	ACT	S=Atr=P



Stoll & Bickel 2013

Data: ACQDIV bank



New Corpora:

Qaqet
Nungon
Pijanjatjara
Ku Waru
Tuatschin
English
Shipibo
etc.

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Turkish	Indonesian	Inuktitut	Sesotho	Dene
Japanese	Yucatec	Chintang	Russian	Cree

Stoll & Bickel 2013

ACQDIV core team

Steven Moran

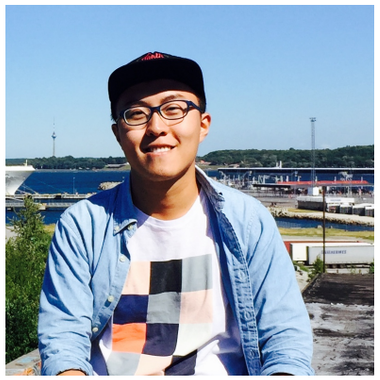
Damian Blasi

Dagmar Jung

Nick Lester

Géraldine Walther

Jekaterina Mazara



Guanghao You

Claudia Cathomas

André Müller

Anna Jansco

Andreas Gerster

Alexandra Bosshard

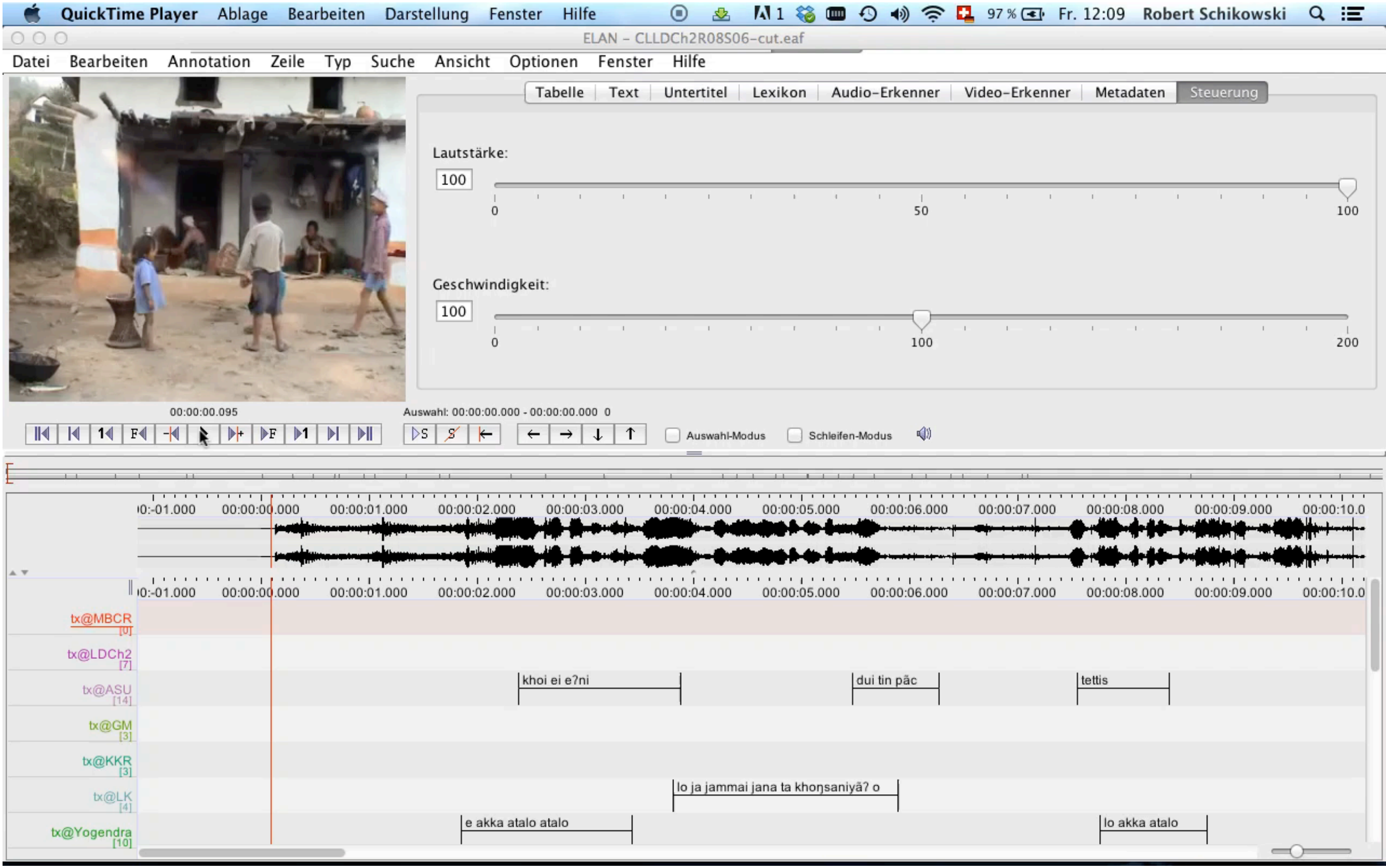
Ruben Moegel

Katherina Habel

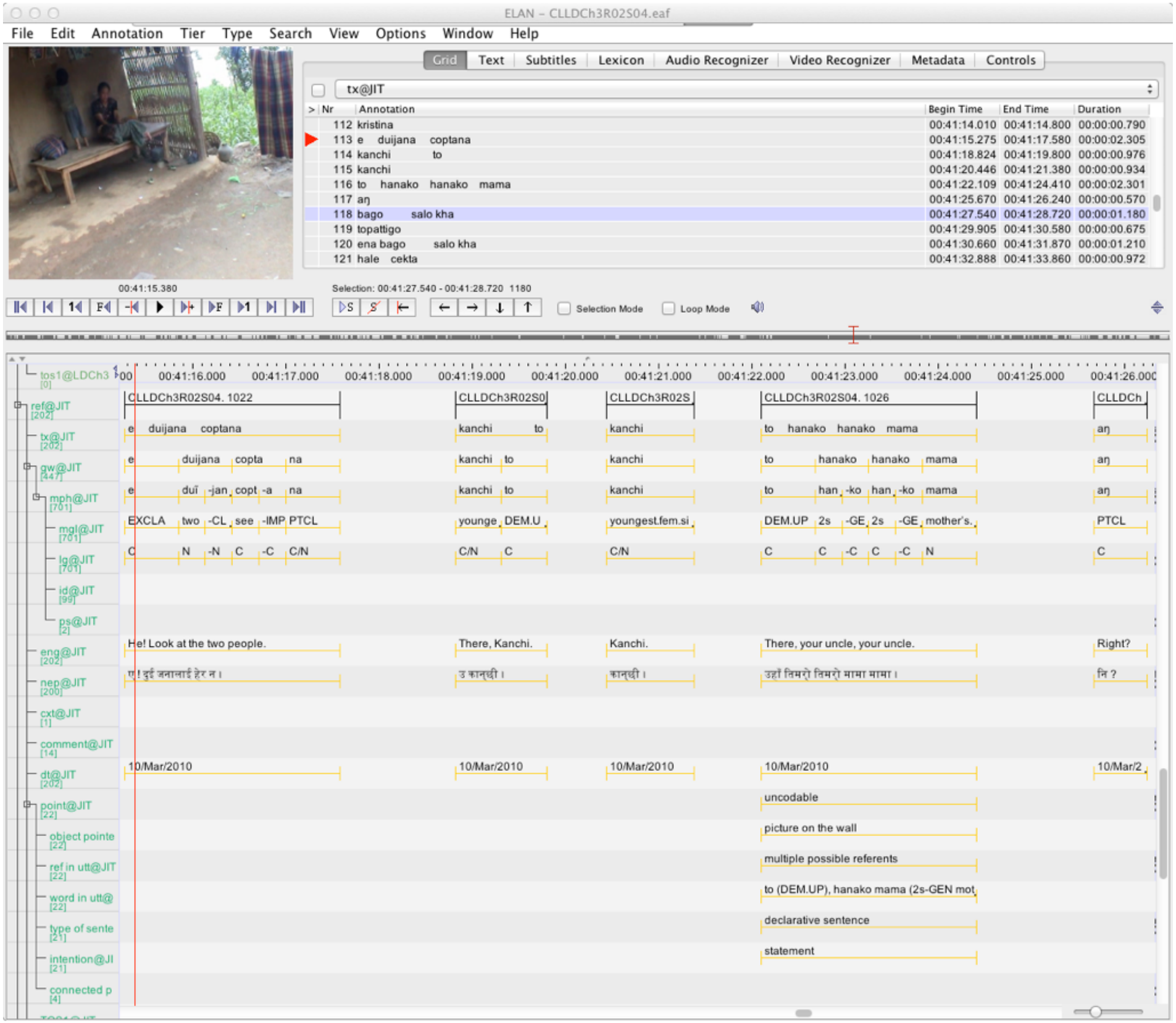
Areezoo Zandy

FNSNF
Swiss National Science Foundation

erc



(www.ivs.uzh.ch/clrp)



(www.ivs.uzh.ch/clrp)

ACQDIV bank

language	corpus	session_id	utterance_id	word_id	morpheme_id	word	pos_word_stem	morpheme	pos	gloss	morpheme_language	speaker_label	age_in_days	gender	role	macrorole	languages_spoken	wa
Chintang	Chintang	1	1	1	1	habiniŋ	V	hap	V	cry	Chintang	GKR	14600	Female	None	Adult	nep bap ctn	NA
Chintang	Chintang	1	1	1	2	habiniŋ	V	i	sfx	1/2pS/P	Chintang	GKR	14600	Female	None	Adult	nep bap ctn	NA
Chintang	Chintang	1	1	1	3	habiniŋ	V	niŋ	sfx	NEG	Chintang	GKR	14600	Female	None	Adult	nep bap ctn	NA
Cree	Cree	478	393608	987673	1600151	awân	PRODEM	NA	PRODEM	who	Cree	ADU	12842	Female	Speaker	Adult	crl	NA
Indonesian	Indonesian	503	410983	1023914	1616384	da	stem	da	stem	bye	Indonesian	MOT	NA	Female	Mother	Adult	xmm ind	NA
Indonesian	Indonesian	503	410983	1023916	1616386	???	???	???	???	???	Indonesian	MOT	NA	Female	Mother	Adult	xmm ind	NA
Indonesian	Indonesian	503	410983	1023917	1616387	Tante	stem	Tante	stem	aunt	Indonesian	MOT	NA	Female	Mother	Adult	xmm ind	NA
Inuktitut	Inuktitut	1500	1326526	3516259	4381281	Naa	PTCL	no	PTCL	no	Inuktitut	ALI	916	Male	Target_Child	Target_Child	ike	NA
Japanese	Japanese_Miyata	1577	1373193	3589311	4419047	issho	N	issho	N	together	Japanese	AMO	NA	Female	Speaker	Adult	und	NA
Japanese	Japanese_Miyata	1577	1373196	3589314	4419048	pikapika	IDEOPH	pikapika	IDEOPH	???	Japanese	AMO	NA	Female	Speaker	Adult	und	NA
Japanese	Japanese_Miyata	1577	1373198	3589317	4419050	ne	PTCL	ne	PTCL	TAG	Japanese	AMO	NA	Female	Speaker	Adult	und	NA
Russian	Russian	1939	1810004	4783322	5470915	oj	INTJ	oj	INTJ	INTERJ	Russian	ALJ	625	Unspecified	Target_Child	Target_Child	NA	NA
Russian	Russian	1939	1810005	4783323	5470916	???	???	???	???	???	Russian	ALJ	625	Unspecified	Target_Child	Target_Child	NA	NA
Russian	Russian	1939	1810006	4783324	5470917	Alja	N	alja	N	M:SG:NOM:AN	Russian	LEN	NA	Female	Aunt	Adult	NA	NA
Russian	Russian	1939	1810006	4783325	5470918	ne	PTCL	ne	PTCL	PCL	Russian	LEN	NA	Female	Aunt	Adult	NA	NA
Sesotho	Sesotho	2389	2637993	6816355	7494533	ere	V	er	V	say	Sesotho	MHL	NA	Female	Mother	Adult	sme	NA
Sesotho	Sesotho	2389	2637993	6816355	7494534	ere	V	e	sfx	m^i	Sesotho	MHL	NA	Female	Mother	Adult	sme	NA
Sesotho	Sesotho	2389	2637993	6816356	7494535	mphe	V	m	pfx	om1s	Sesotho	MHL	NA	Female	Mother	Adult	sme	NA
Sesotho	Sesotho	2389	2637993	6816356	7494536	mphe	V	ph	V	give	Sesotho	MHL	NA	Female	Mother	Adult	sme	NA
Turkish	Turkish	2667	2858439	NA	7824079	NA	NA	değil	PTCL	???	Turkish	NA	NA	NA	NA	NA	NA	NA
Turkish	Turkish	2667	2858439	NA	7824080	NA	NA	o	PRODEM	???	Turkish	NA	NA	NA	NA	NA	NA	NA
Turkish	Turkish	2667	2858440	7504733	7824081	kamera	N	kamera	N	???	Turkish	MOT	NA	Female	Mother	Adult	xmm ind	NA
Yucatec	Yucatec	2891	3108071	8185947	8123413	Lorena-e'	N	Lorena	N	???	Yucatec	MAR	21486	Male	Speaker	Adult	jav ind	NA
Yucatec	Yucatec	2891	3108071	8185947	8123414	Lorena-e'	N	e'	sfx	TOP	Yucatec	MAR	21486	Male	Speaker	Adult	jav ind	NA
Yucatec	Yucatec	2891	3108071	8185948	8123415	táan	???	táan	???	???	Yucatec	MAR	21486	Male	Speaker	Adult	jav ind	NA
Yucatec	Yucatec	2891	3108071	8185949	8123416	u	???	u	???	???	Yucatec	MAR	21486	Male	Speaker	Adult	jav ind	NA

Moran, Schikowski, Pajović, Hysi & Stoll, 2016

Steven Moran

Nick Lester

Study 2: Variation sets

Moran, Lester, Heath, Küntay Pfeiler, Allen & Stoll 2019, Cognition

Hypothesis

There are patterns in the input that make learning and generalisation possible.

Learning in interaction

Two hypothetical language learning situations.

Alien message #1	**Alien message #2**
(1a) kedmalburafuloropesai	(1b) kedmalburafuloropesai
(2a) gianaber	(2b) rafuloro
(3a) manadukbiunel	(3b) manaloropesai
(4a) kiciorudanamjeisulcaz	(4b) kedmalbumanaloropesai

Onnis et al. 2008

Variation sets in the input

*MOT: und jetz rein in den schuh.
*MOT: und die schuhe?
*MOT: weg .
*MOT: so zumachen.
*MOT: eine schleife.
*MOT: oh 'n grossen schuh.
*MOT: bum.
*MOT: warte.
*MOT: du kriegst gleich dein'n schuh wieder.
*CHI: Mama.
*MOT: nein nich?
*MOT: wieder zu?
*CHI: zu.
*MOT: zu?
*CHI: &ja schuh &em.
*MOT: schuh?
*MOT: ein schuh?
*MOT: die puppe is zu klein für den grossen schuh?

Childes, German/Szagun/CI/Adriane/030520.cha

Variation sets in the input

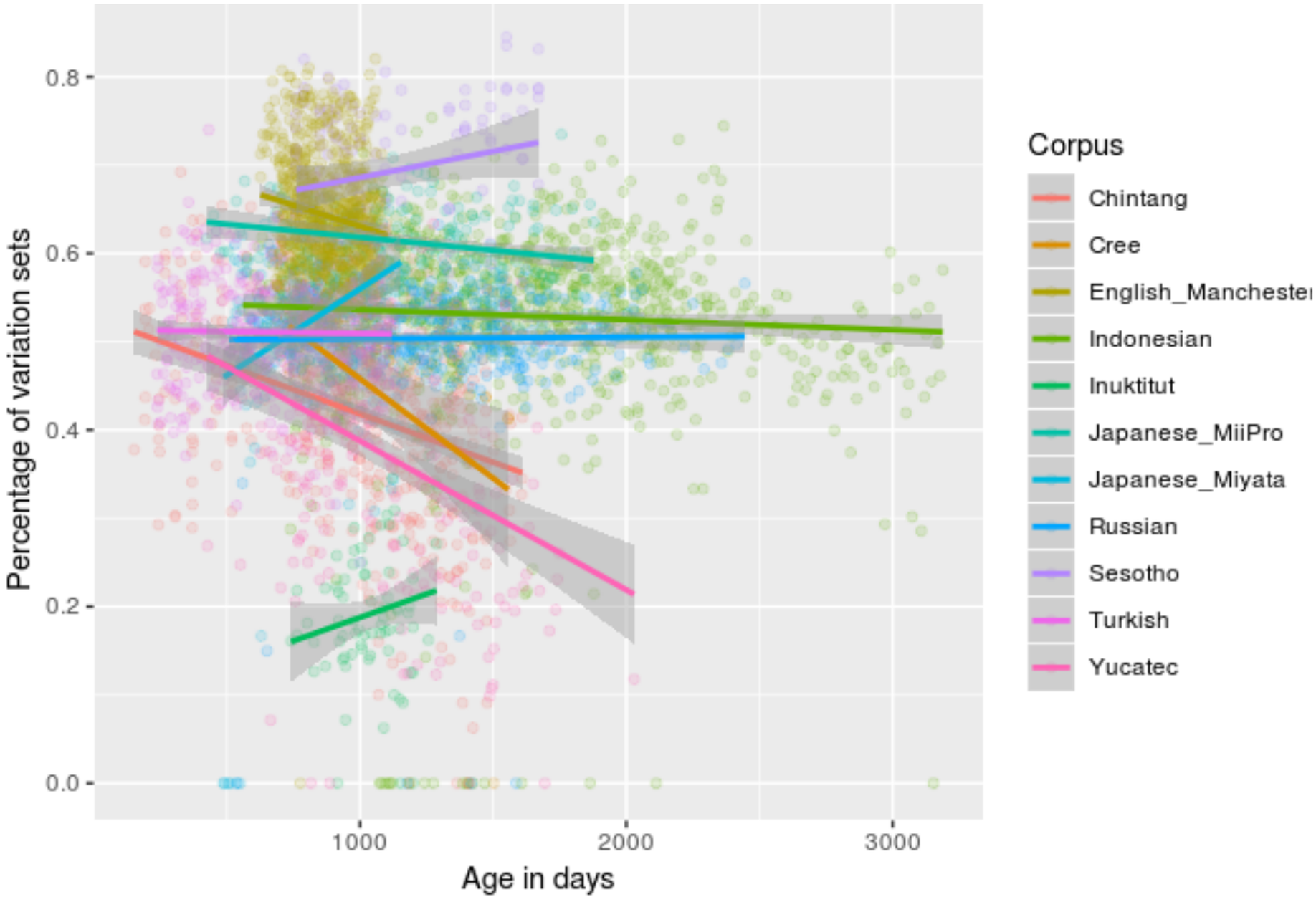
- Repetitive units in interactional units

- 17% - 30% of verbs and nouns in variation sets in child-directed speech

Küntay & Slobin 1996, Wiren et al. 2016, Waterfall 2006, Brodsky et al. 2007, Grigonyte & Björkenstamm 2016

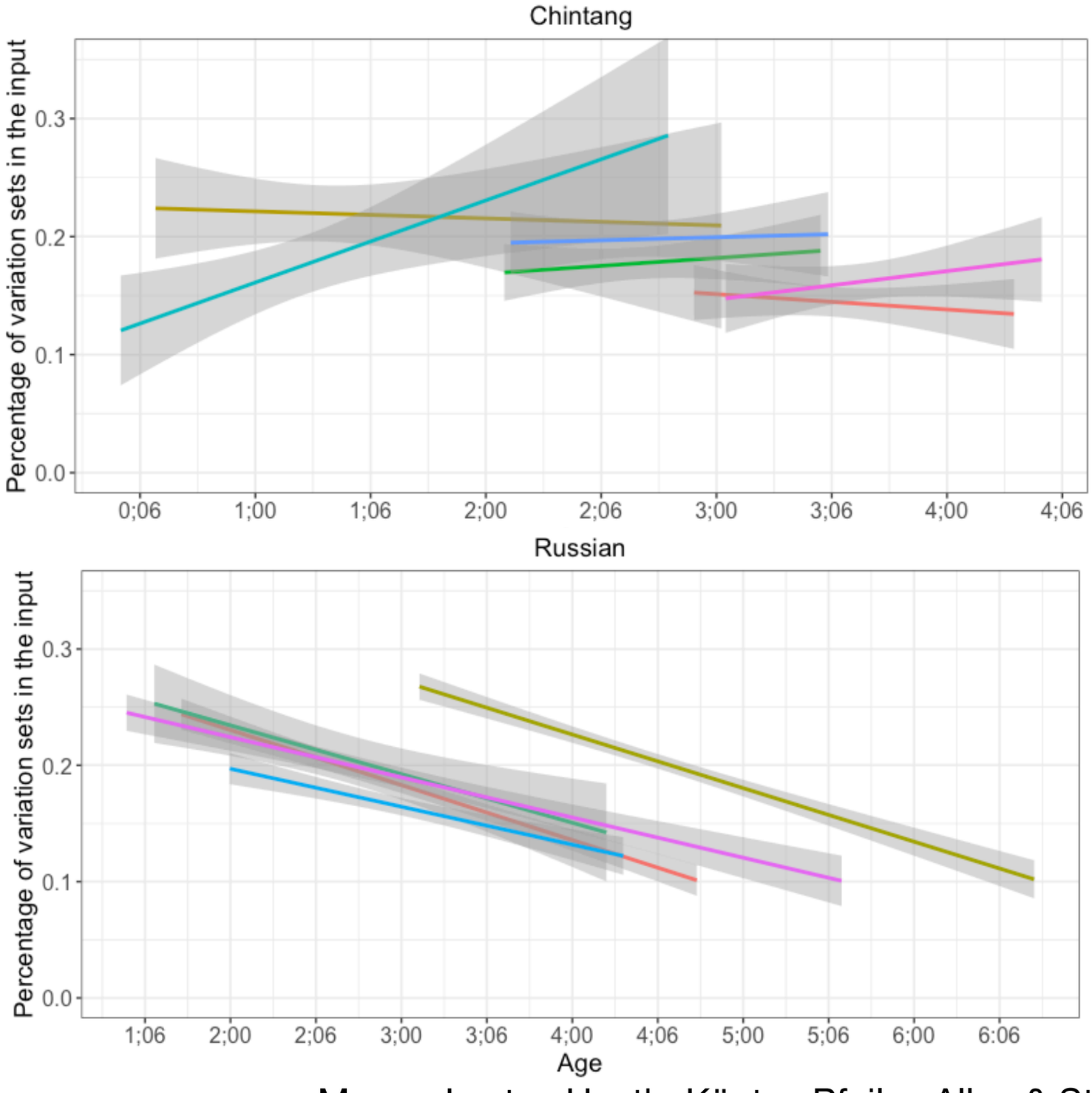
Are *variation sets* in input universally reliable patterns and how do they develop as a function of the age of the child?

Variation sets



Moran, Lester, Heath, Küntay Pfeiler, Allen & Stoll 2019

Variation sets



Moran, Lester, Heath, Küntay Pfeiler, Allen & Stoll 2019

Conclusion variation sets

Found in all of our maximally diverse languages
Language specific patterns in the development over time, not necessarily decreasing

Steven Moran

Study 3: Frames to detect nouns and verbs

Frames

I	like	you
I	hate	you
I	hear	you
I	love	you
I	and	you
I	not	you

Moran, Blasi, Schikowski, Küntay, Pfeiler, Allen & Stoll , Cognition 2018

Frames

Frame you__it
put (28), want (15), do (10), see (7), take (6), turn (5), taking (5), said (5), sure (4), lost (4), like (4), leave (4), got (4), find (4), throw (3), threw (3), think (3), sing (3), reach (3), picked (3), get (3), dropped (3), seen (2), lose (2), know (2), knocked (2), hold (2), help (2), had (2), gave (2), found (2), fit (2), enjoy (2), eat (2), chose (2), catch (2), with (1), wind (1), wear (1), use (1), took (1), told (1), throwing (1), stick (1), share (1), sang (1), roll (1), ride (1), recognize (1), reading (1), ran (1), pulled (1), pull (1), press (1), pouring (1), pick (1), on (1), need (1), move (1), manage (1), make (1), load (1), liked (1), lift (1), licking (1), let (1), left (1), hit (1), hear (1), give (1), flapped (1), fix (1), finished (1), drop (1), driving (1), done (1), did (1), cut (1), crashed (1), change (1), calling (1), bring (1), break (1), because (1), banged (1)

Frame *a-* '2S/A' __ *-o* '3P'

a. *theke a-**kʰind**-o-ko?*
why 2S/A-pull-3P-IND.NPST
'Why do you pull it?'

b. *hũi kãhili a-**nept**-o*
DEM third.daughter 2S/A-step.on-3P
'Kaĩli, you might step on it!'

c. *a-**hid**-o-ko?*
2S/A-watch-3P-IND.NPST
'Can you look after her?'

Frame *mai-* 'NEG' __ *-th* 'NEG'

a. *la mo mai-**soŋ**-th-a*
INTERJ DEM.DOWN NEG-move-NEG-IMP
'Don't move down there!'

b. *e mai-**khaŋ**-th-o-s-e ni*
OK NEG-see-NEG-3P-PRF-IND.PST EMPH
'Yeah, s/he didn't see it.'

c. *mo sotaŋma-ce phidaŋ u-mai-**let**-th-a-ŋs-a-kha*
DEM.DOWN S.-ns ginger 3nsS/A-NEG-plant-NEG-PST-PRF-PST-NMLZ
'The Sotangma people have not yet planted ginger down in the field.'

Are *frames* in the input universally reliable patterns that could help in categorizing parts of speech?

Frames in ACQDIV bank

Word frames: no

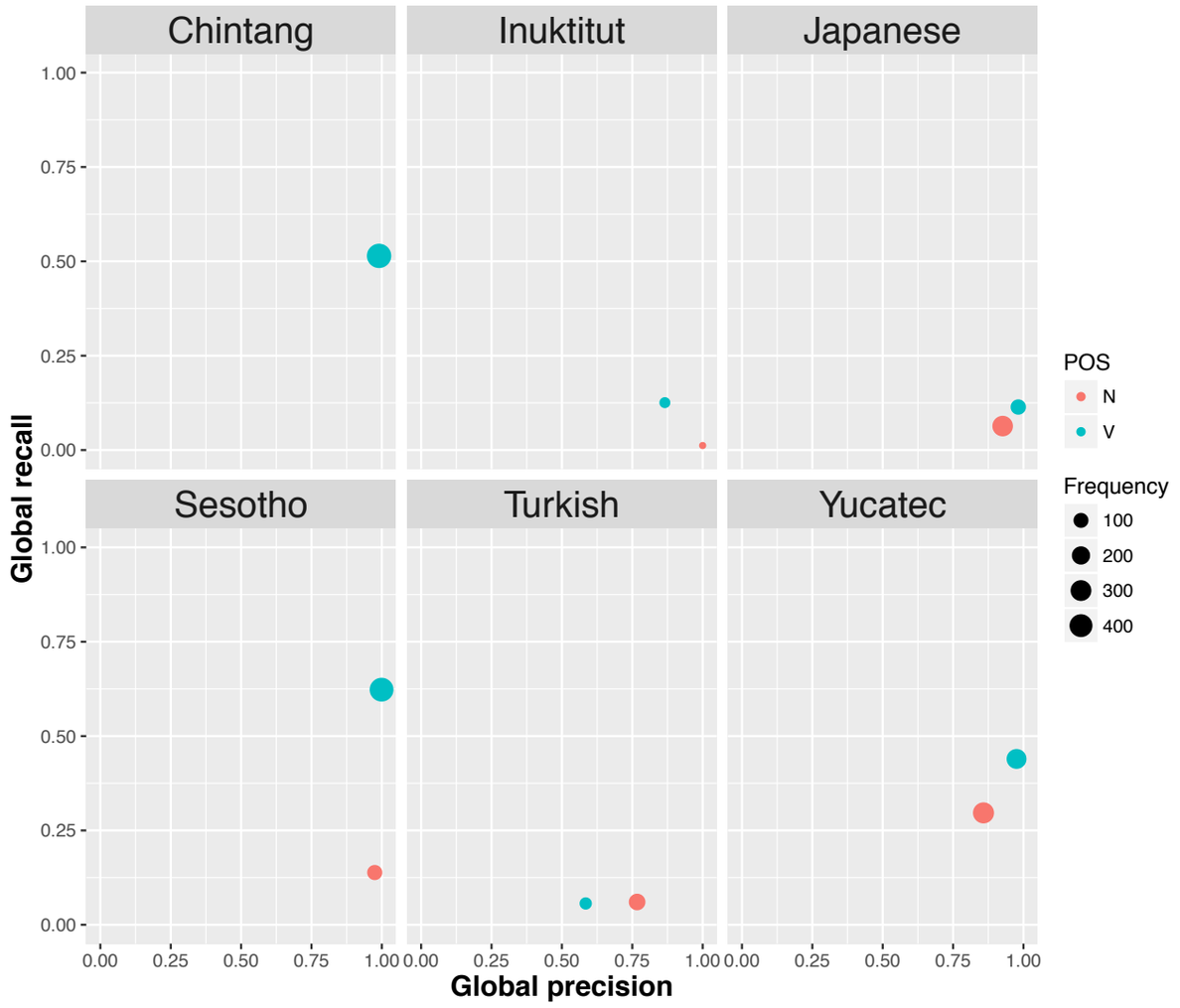
	Accuracy	SD	Completeness	SD	Frames	Min	Max
Chintang	0.57	0.24	0.04	0.02	33	90	2720
Inuktitut	0.98	0.11	0.03	0.01	37	2	3
Japanese	0.82	0.21	0.02	0.02	97	67	915
Russian	0.44	0.22	0.04	0.03	48	234	1485
Sesotho	0.83	0.23	0.01	0.01	107	8	163
Turkish	0.62	0.20	0.08	0.08	15	34	318
Yucatec	0.78	0.28	0.01	0.01	133	3	41

Morpheme frames: yes

	Accuracy	SD	Completeness	SD	Frames	Min	Max
Chintang	0.95	0.09	0.08	0.07	60	517	7940
Inuktitut	0.93	0.16	0.02	0.01	100	5	43
Japanese	0.98	0.04	0.02	0.03	187	83	1943
Sesotho	0.97	0.12	0.04	0.04	88	66	1358
Turkish	0.88	0.17	0.01	0.01	835	21	1000
Yucatec	0.90	0.18	0.01	0.02	153	20	584

Moran, Blasi, Schikowski, Küntay, Pfeiler, Allen & Stoll , Cognition 2018

Categorization of nouns and verbs via frames



Moran, Blasi, Schikowski, Küntay, Pfeiler, Allen & Stoll , Cognition 2018

Conclusion frames:

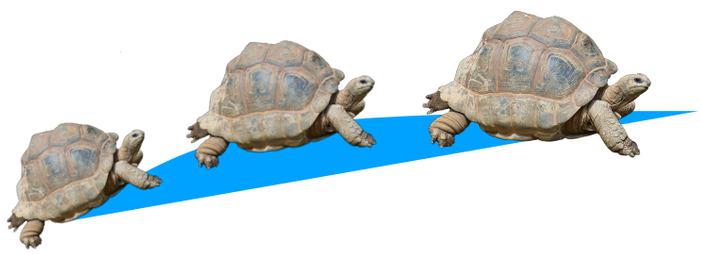
Nouns and verbs can be categorized via repetitive patterns in their morphological structure

Nick Lester

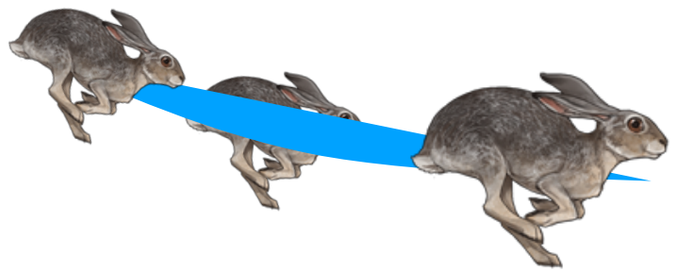
Study 4:

Temporal cues to detect nouns and verbs

Temporal cues before nouns and verbs



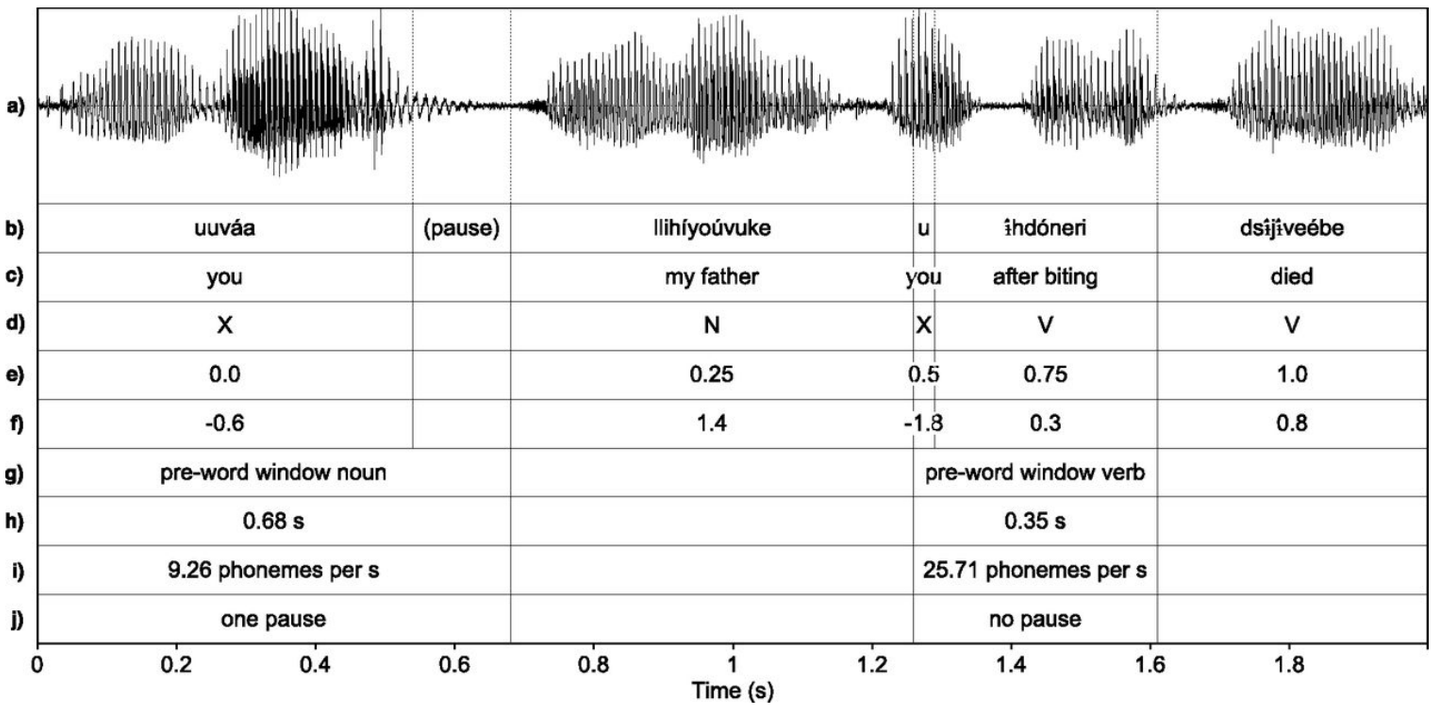
nouns

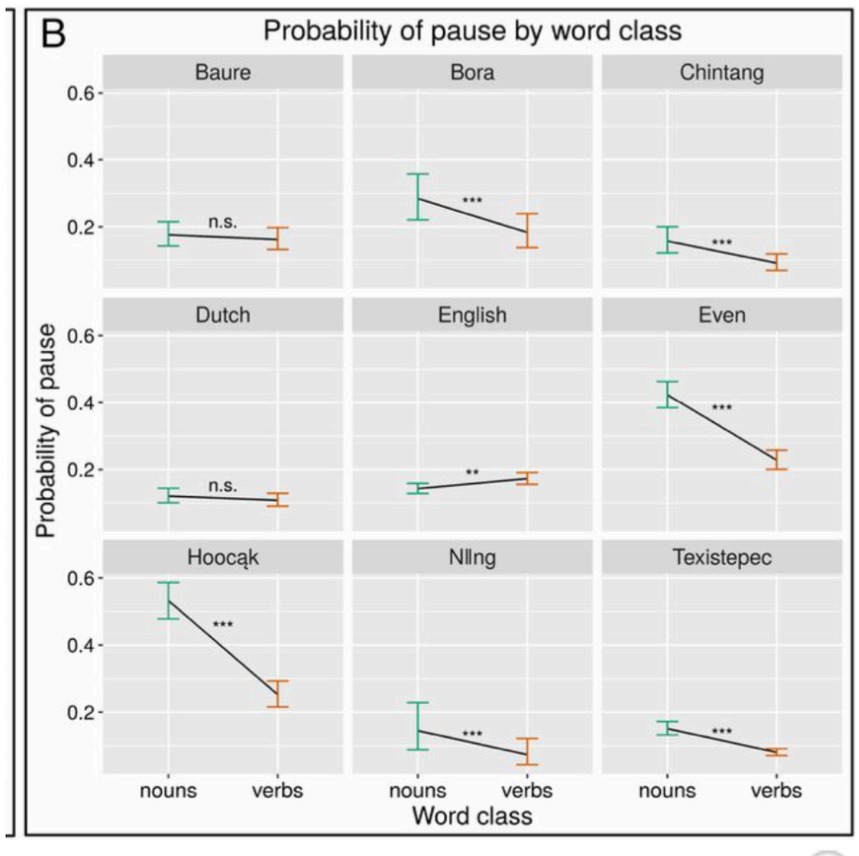


verbs

Seifart … Bickel, PNAS 2018

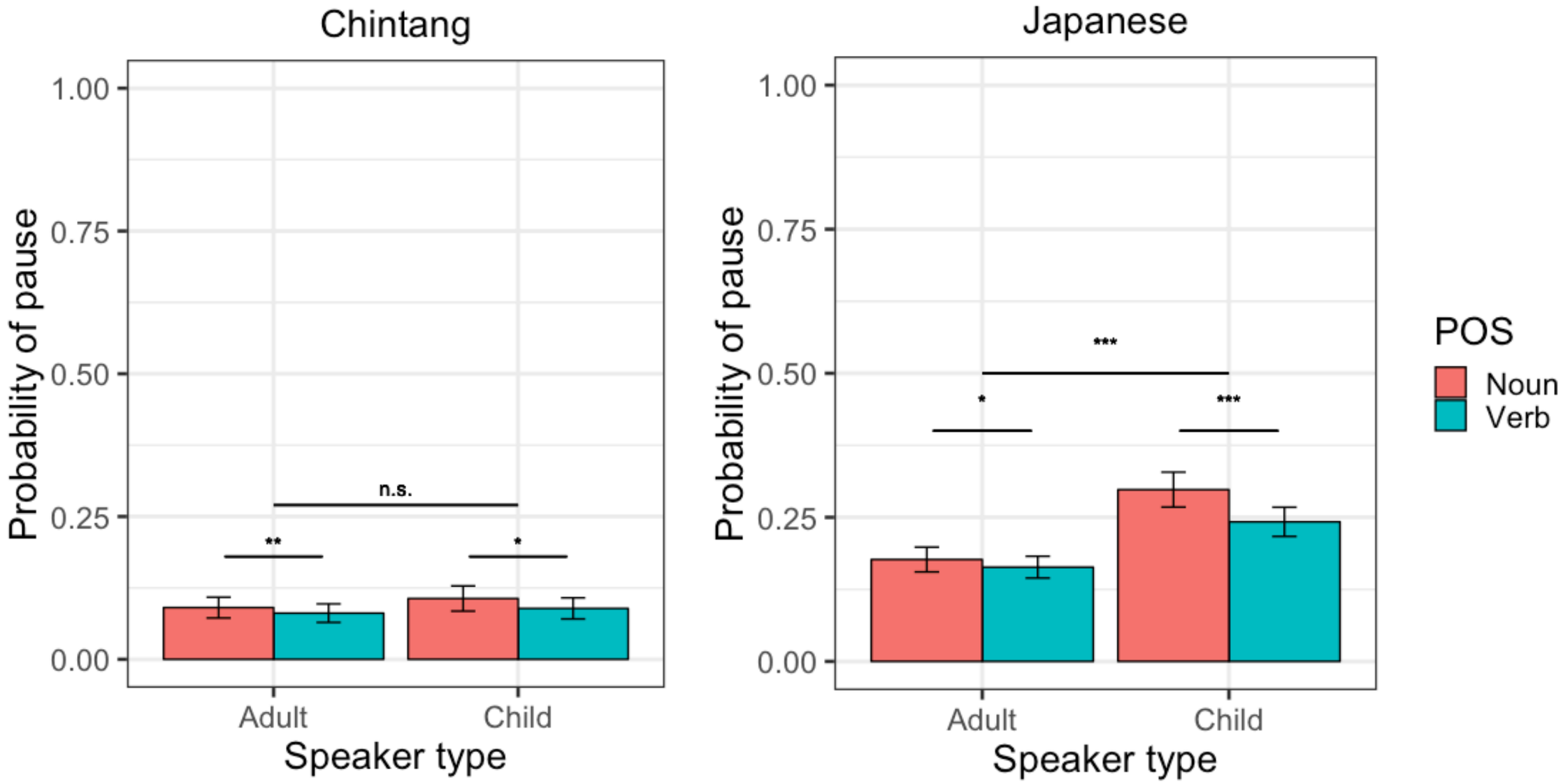
Temporal cues before nouns and verbs: adults





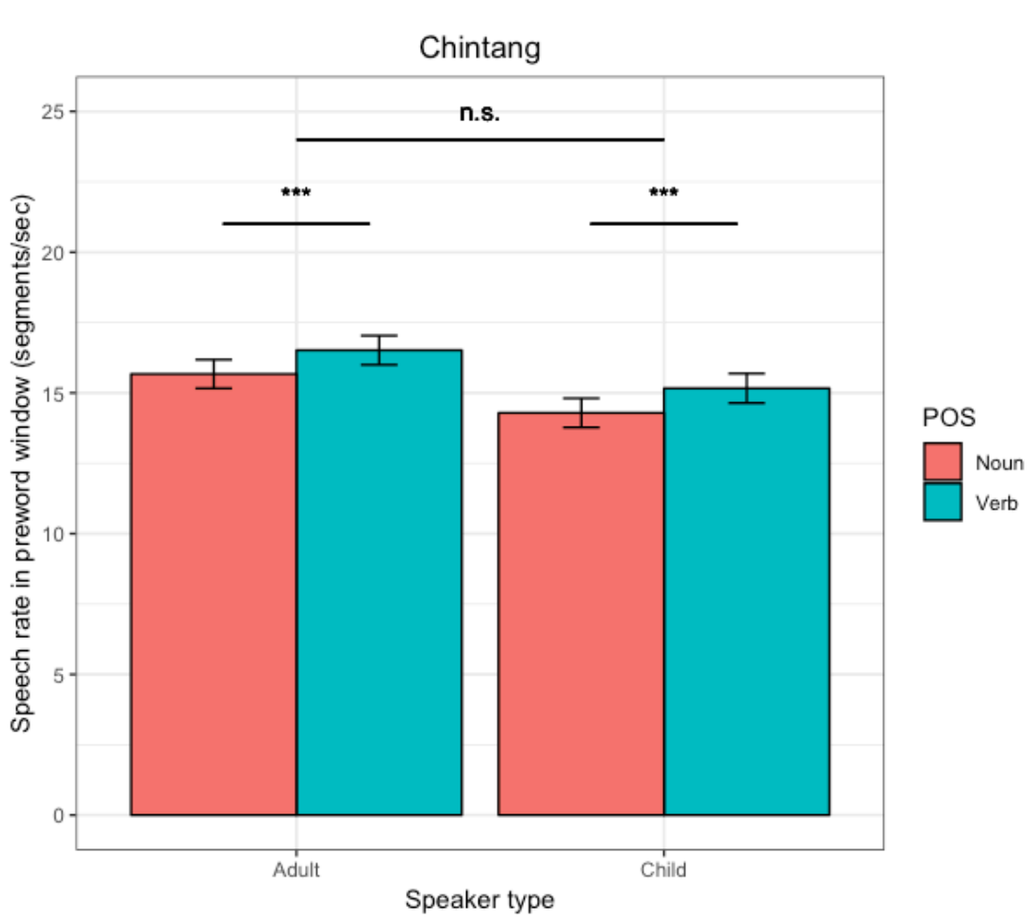
Seifart … Bickel, PNAS 2018

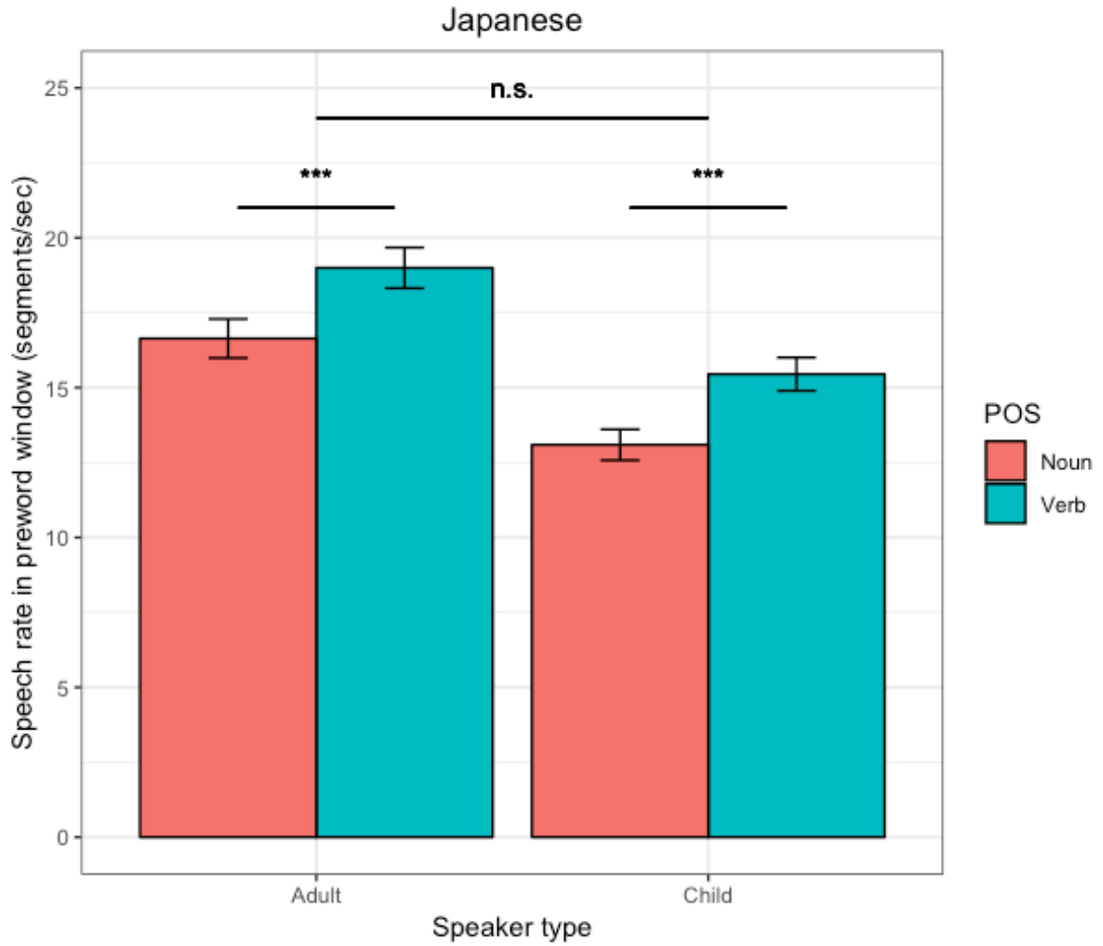
Temporal cues for nouns and verbs: pauses in child-surrounding speech



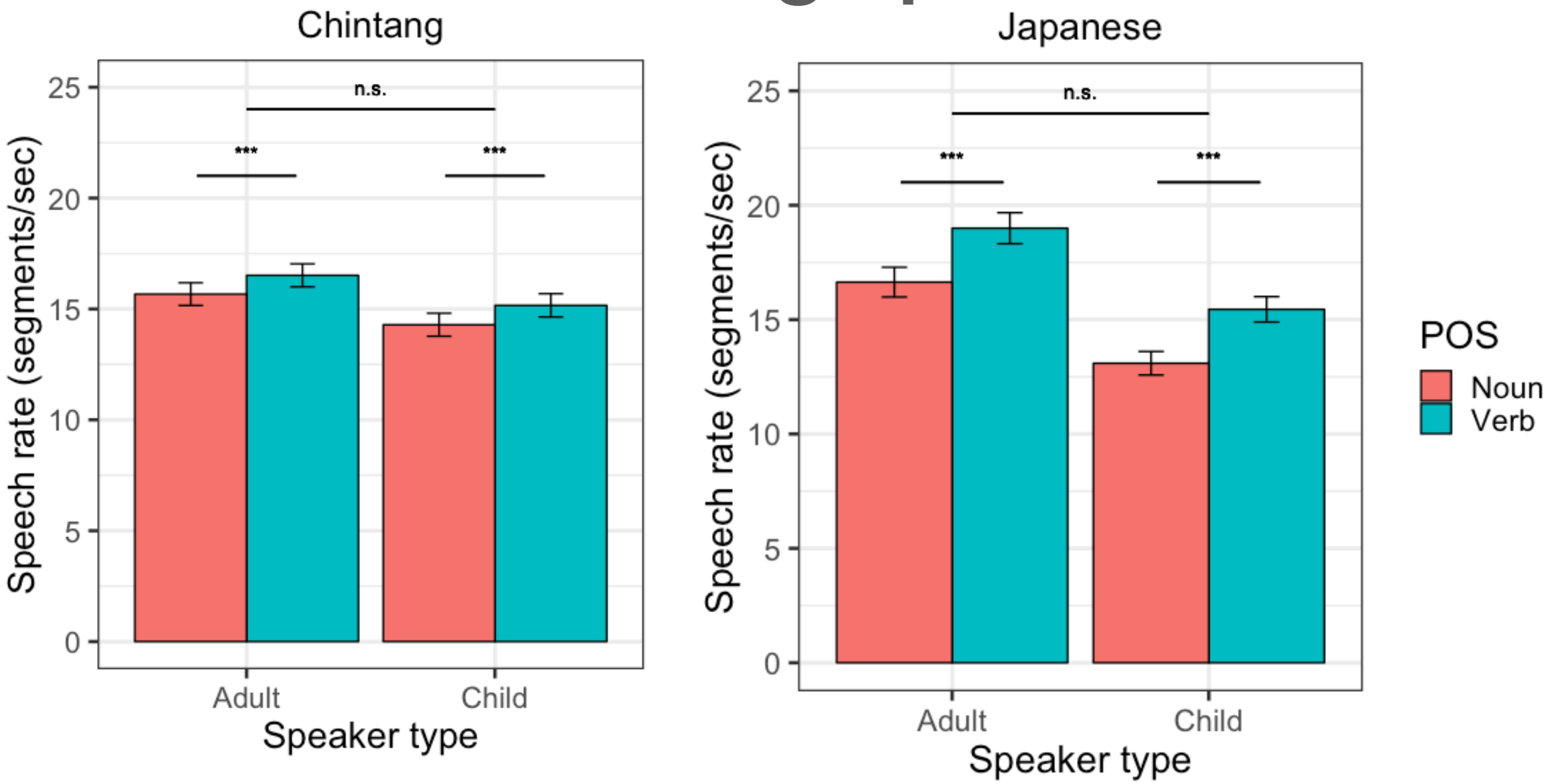
Lester, Bickel & Stoll in prep.

Sprechgeschwindigkeit vor Nomen oder Verben





Temporal cues for nouns and verbs: speech rate in child-surrounding speech



Lester, Bickel & Stoll in prep.

Conclusion temporal cues:

Both pauses before nouns and verbs and the temporal duration of these units can serve as cues for syntactic categorization

To conclude:

- The input is not chaotic and there are quite a number of patterns learners can rely on
- There are striking similarities in the information distribution in child-directed speech in maximally diverse languages
- Some cross-linguistic differences in the patterns
- We are only at the beginning at understanding the computational power underlying language development
- Children can apply distributional learning and make use of statistical patterns provided by the input

Thank you very much for your interest!