



Statistical learning allows children to efficiently acquire even the most complex languages

**Sabine Stoll
Jekaterina Mazara
Damian Blasi
Balthasar Bickel**

How can we study language development?

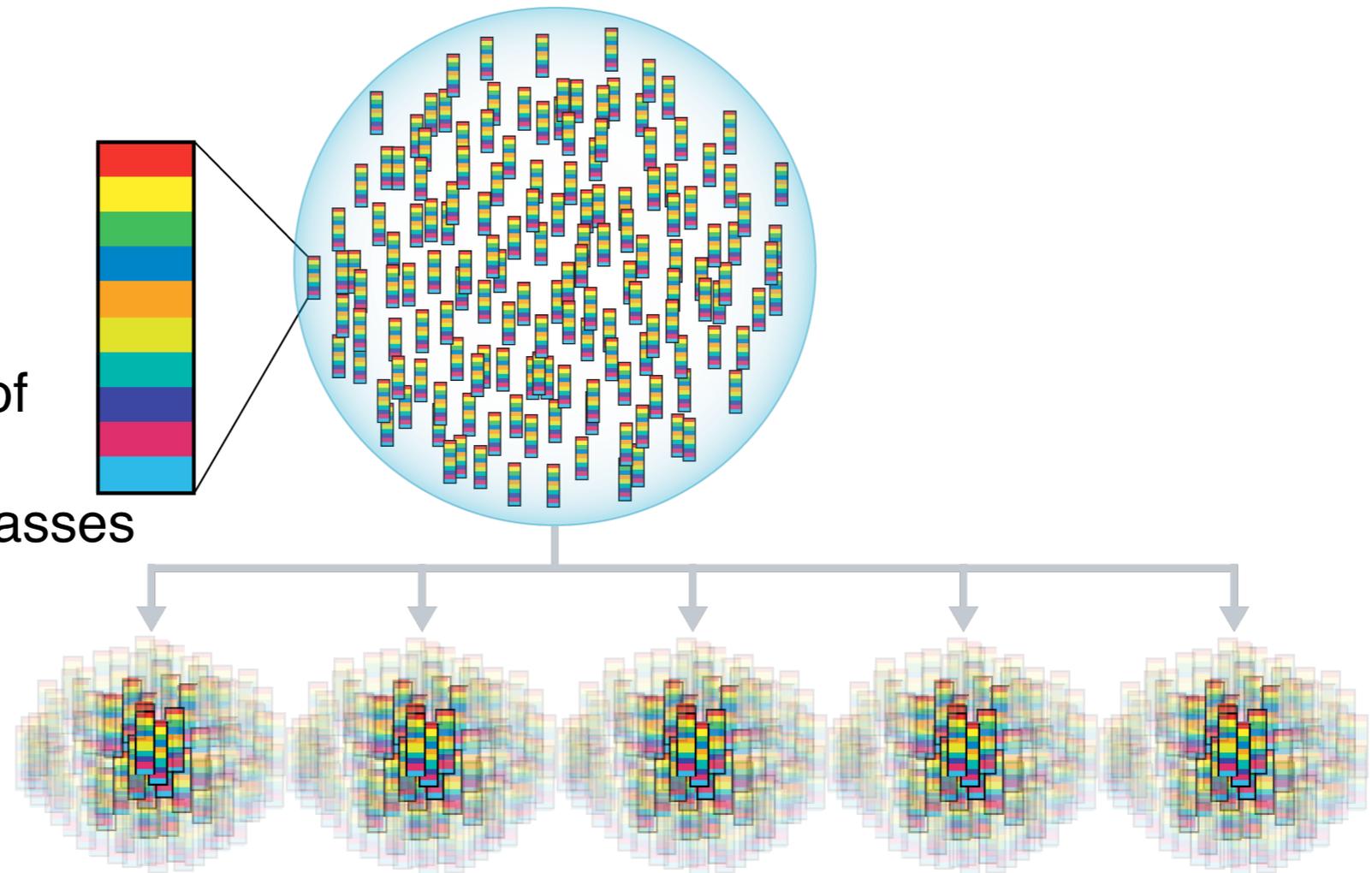
- **main question:** how do children build up the grammatical structure of their languages?
- **our focus:** how do they build up verbal morphology and how can we measure this process?
- **goal:** search for general, potentially universal learning mechanisms
- **complication:** task varies considerably cross-linguistically because of different degrees of complexity in morphology
 - number of forms
 - combinability and variability of expression of these elements (e.g. exponence patterns)

Key idea: Maximum-Diversity Approach

- sample from maximally diverse languages

Language features

word order
synthesis
exponence
case marking
inflectional compactness of categories
existence of inflectional classes
...



Satellites
(sign language)
Kata Kolok
Tzeltal

Cluster 1
Turkish
Japanese

Cluster 2
Indonesian
Yucatec

Cluster 3
Inuktitut
Chintang

Cluster 4
Sesotho
Russian
English

Cluster 5
Dene
Cree



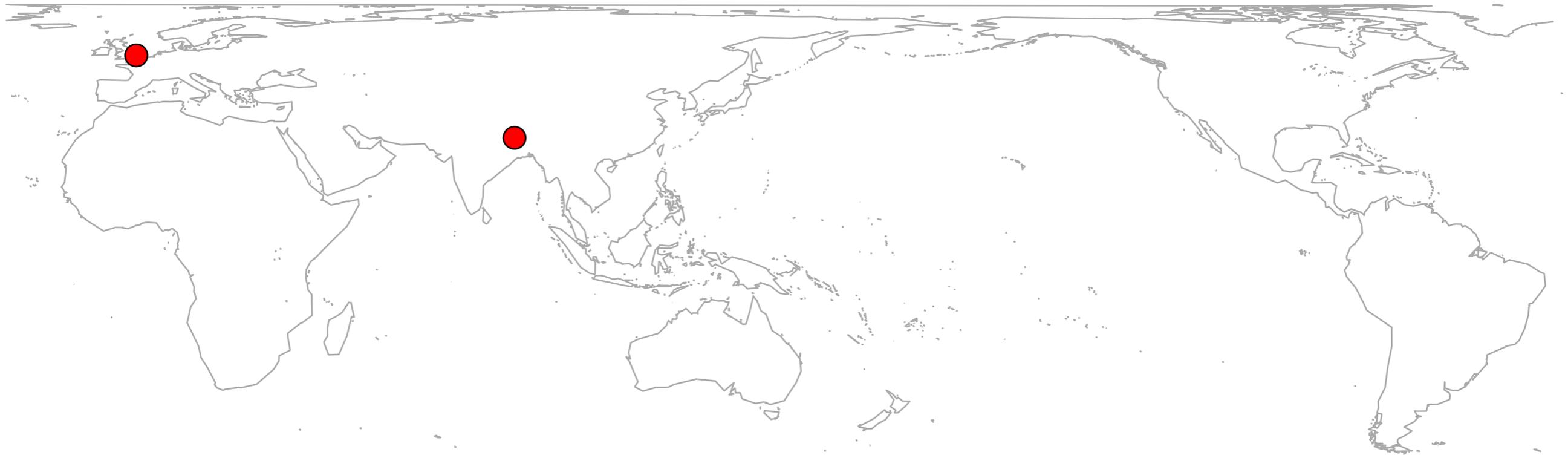
University of
Zurich ^{UZH}

Stoll & Bickel 2013



Data

- longitudinal corpora of English and Chintang (Sino-Tibetan, Eastern Nepal)
 - Chintang: polysynthetic
 - English: very little morphology



Documentation of Chintang

Chintang Language Research Program
(<http://www.clrp.uzh.ch>)



University of
Zurich ^{UZH}

Chintang Language Research Program
(<http://www.clrp.uzh.ch>)



Documentation of Chintang

Linguistics:

Grammar
Discourse
Lexicon

How is the language
structured?

Ethnography:

Rituals and religion
Social structure
Everyday life

How is the language
used?

Psycholinguistics:

Acquisition strategies
Multilingualism

How is the language
learned?



Balthasar
Bickel

Robert
Schikowski

Novel K.
Rai

Martin
Gaenszle

Sabine
Stoll

Elena
Lieven

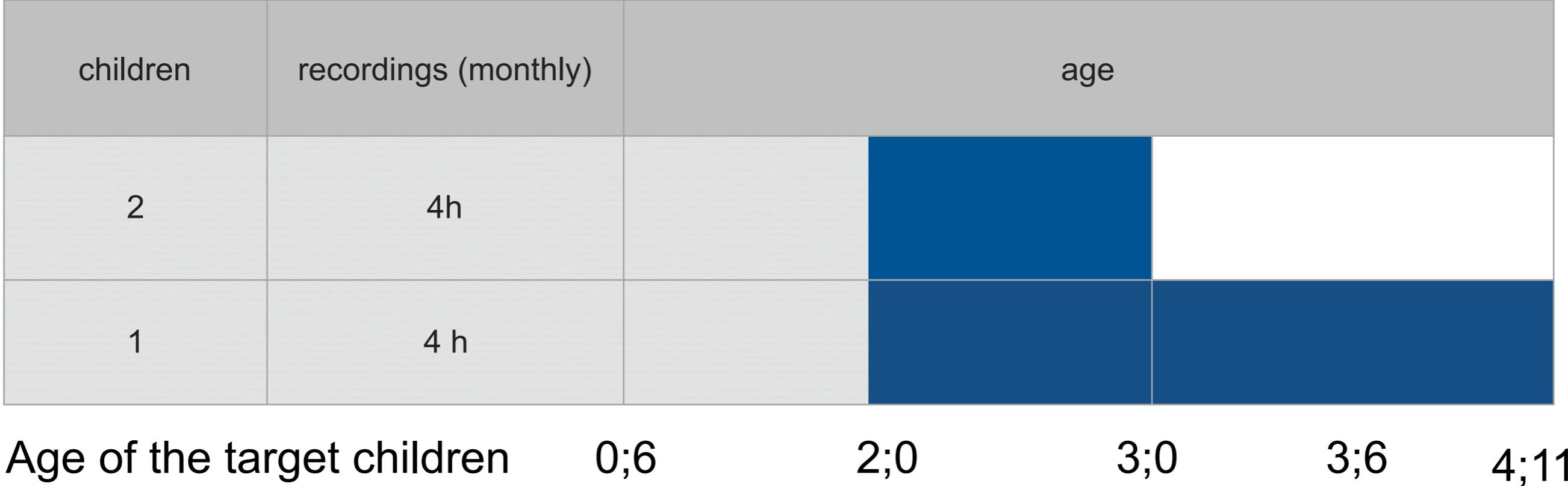


The Chintang Language Acquisition Corpus

- The Chintang Language Corpus

Genre	Transcribed	Transcribed	Translated	Translated # words	Glossed duration	Glossed # words
conversation	232:44:54	1'064'109	232:39:18	1'045'254	207:23:27	903'645
description	3:09:10	20'934	3:06:24	20'617	1:52:56	14'433
narrative	6:15:13	46'044	6:12:12	45'826	5:54:52	42'922
experimental	4:11:31	43'780	3:57:37	33'273	2:48:08	24'110
ritual language	11:15:55	40'634	10:03:34	32'736	6:36:02	18'270
songs	00:57:27	3'264	0:31:00	1'588	0:21:48	1'231
total	258:34:10	1'218'765	256:30:05	1'179'294	225:02:07	1'004'837

The Manchester Language Acquisition Corpus



- high density acquisition corpus, matched to the same number of recordings per child as in the Chintang corpus

Morphological complexity: English vs.

	Number of affixes	Number of unique verb forms
English	3	3
Chintang	148	4745

Morphological complexity: English vs. Chintang

English

Chintang

I go
 you go
 he **goes**
 we go
 you go
 they go
went

	1s	1di	1pi	1de	1pe	2s	2d	2p	3s	3ns	intransitive
1s						tupnaʔā tupnaʔāniŋ tupnehē matupyoknehē	tupnaʔāce tupnaʔāceŋiŋ tupnace matupyoknace	tupnaʔāni tupnaʔāniŋiŋ tupnanihē matupyoknanihē	tubukuj tubukujniŋ tubuhē matupyoktuhē	tubukujcuŋ tubukujcuŋniŋ tubujcihē matupyoktucihē	tupmaʔā tupmaʔāniŋ tubehē matupyoktehē
1di						tupcoko tupcokoniŋ tubace matupyoktace	tubumcum tubumcumniŋ tubumcumhē matupyoktumcumhē	tupceke tupcekeniŋ tubace matupyoktace			
1pi						tubukum tubukumniŋ tubumhē matupyoktumhē	tubiki tubikiniŋ tubihē matupyoktihē				
1de						tupnaʔānciyā tupnaʔānciyāniŋ tupnanciyehē matupyoknanciyehē	tupcokoya tupcokoyaniŋ tubacehē matupyoktacehē	tubumcumma tubumcummaniŋ tubumcummehe matupyoktumcummehe	tupcekeja tupcekejaniŋ tubacehē matupyoktacehē		
1pe						tubukumma tubukummaniŋ tubummehē matupyoktummehe	tubikija tubikijaniŋ tubiehē matupyoktiehē				
2s	atupmaʔā atupmaʔāniŋ atubehē {a-ma}tupyoktehē		{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				atuboko atubokoniŋ atube amatupyokte	atubukuce atubukuceŋiŋ atubuce {a-ma}tupyoktuce	atupno atupnikniŋ atube {a-ma}tupyokte	
2d	atupmaʔānciŋ atupmaʔānciŋniŋ atubajcihē {a-ma}tupyoktaŋcihē							atupcoko atupcokoniŋ atubace amatupyoktace	atubumcum atubumcumniŋ atubumcumhē {a-ma}tupyoktumcumhē	atupceke atupcekeniŋ atubace {a-ma}tupyoktace	
2p	atupmaʔāniŋ atupmaʔāniŋiŋ atubajnihē {a-ma}tupyoktaŋnihē							atubukum atubukumniŋ atubumhē amatupyoktumhē		atubiki atubikiniŋ atubihē {a-ma}tupyoktihē	
3s	utupmaʔā utupmaʔāniŋ utubehē {u-ma}tupyoktehē	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 natubihē {na-ma}tupyoktihē	tuboko tubokoniŋ tube matupyokte	tubukuce tubukuceŋiŋ tubuce matupyoktuce	tupno tupnikniŋ tube matupyokte
3d	utupmaʔānciŋ utupmaʔānciŋniŋ utubajcihē {u-ma}tupyoktaŋcihē							utupcoko utupcokoniŋ utubace {u-ma}tupyoktace	utubukuce utubukuceŋiŋ utubuce {u-ma}tupyoktuce	utupceke utupcekeniŋ utubace {u-ma}tupyoktace	
3p	utupmaʔāniŋ utupmaʔāniŋiŋ utubajnihē {u-ma}tupyoktaŋnihē							utuboko utubokoniŋ utube {u-ma}tupyokte		utupno utupnikniŋ utube {u-ma}tupyokte	

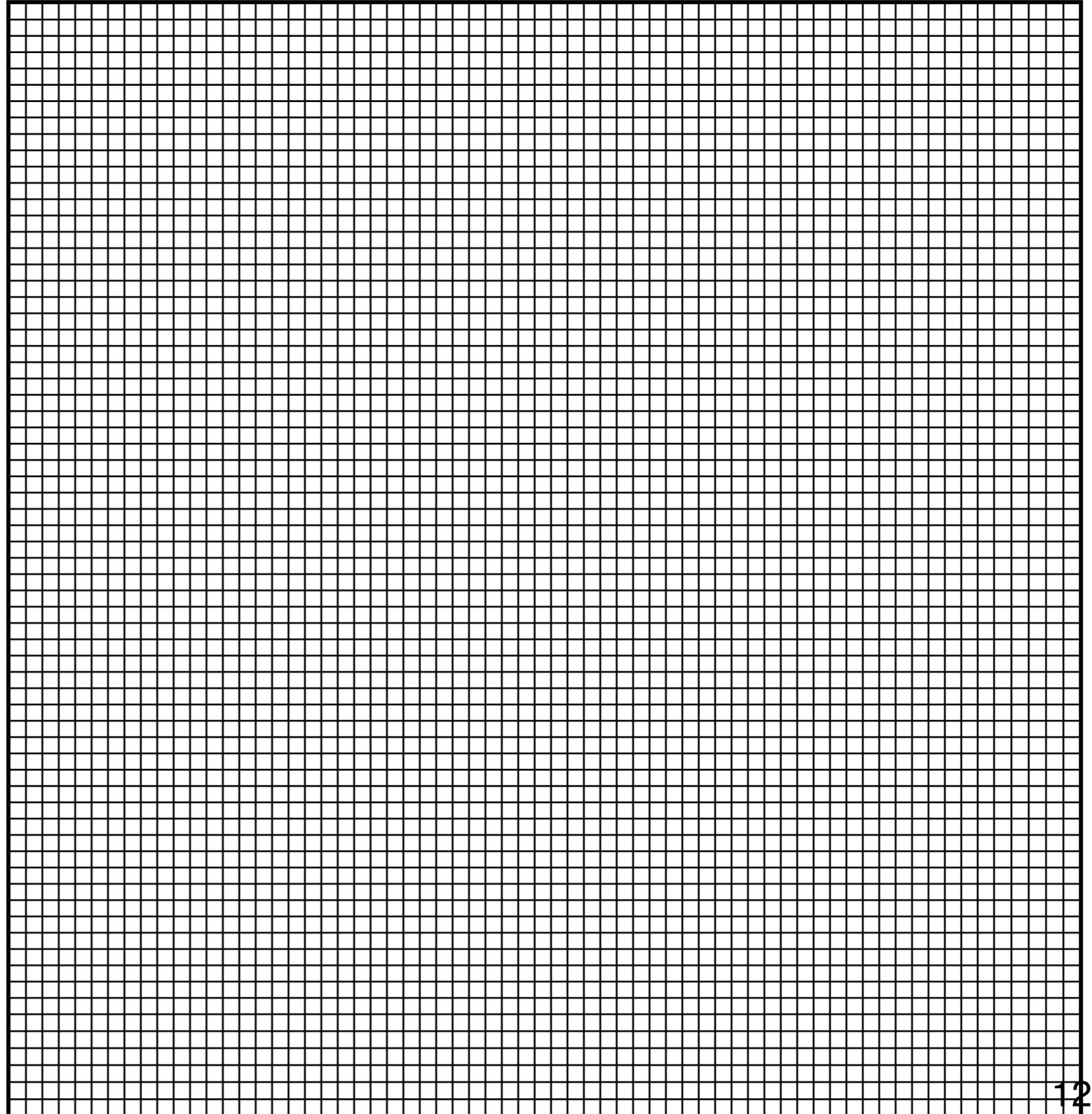
jo-go-yaŋ

na-khutt-i-ca-i-hatt-i-bir-i.

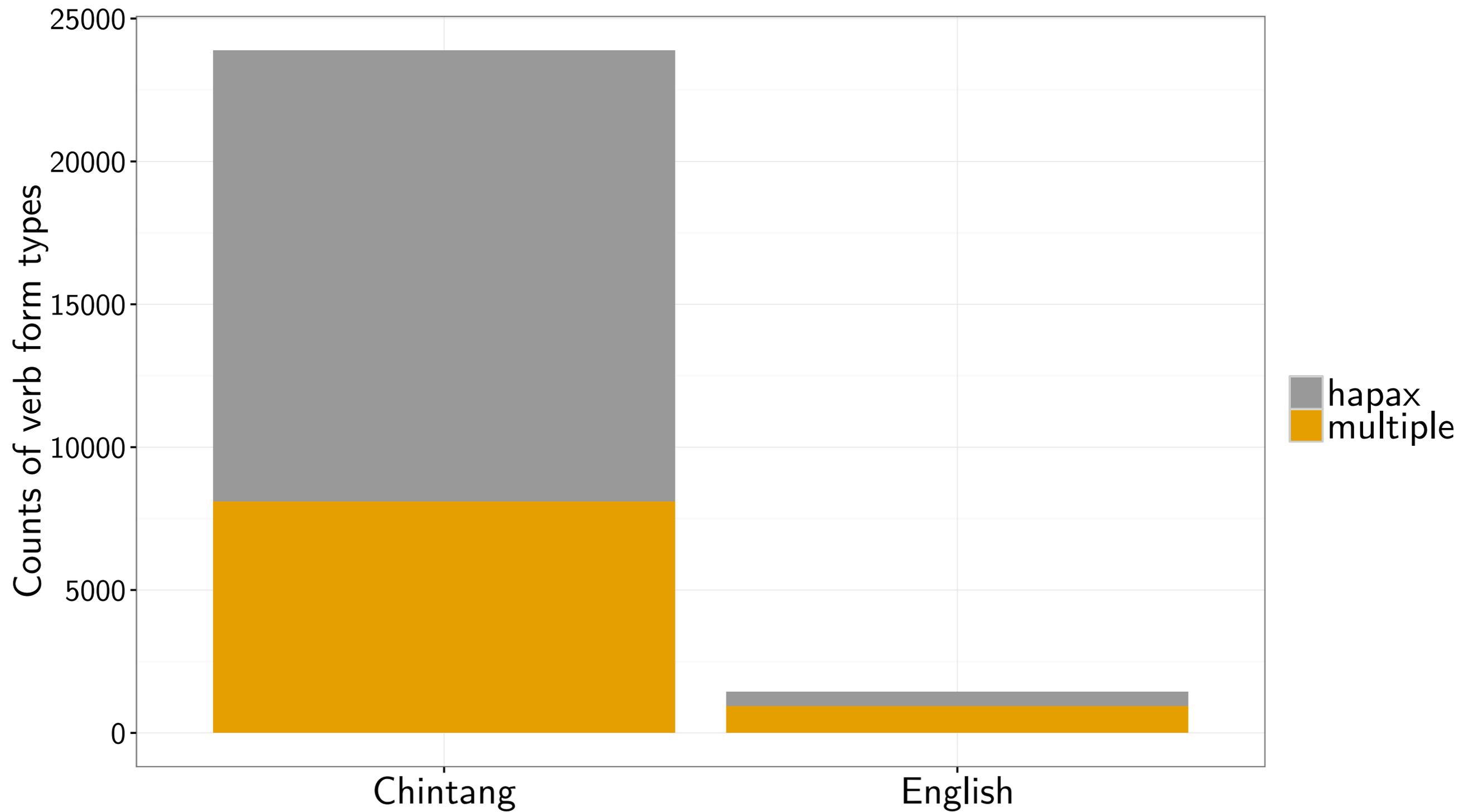
whatever-NMLZ-ADD 3[s]>2-steal-2pO-V2:eat-2pO-V2:move.away.TR-2pO-V2:do.for-[SBJV.]2pO

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

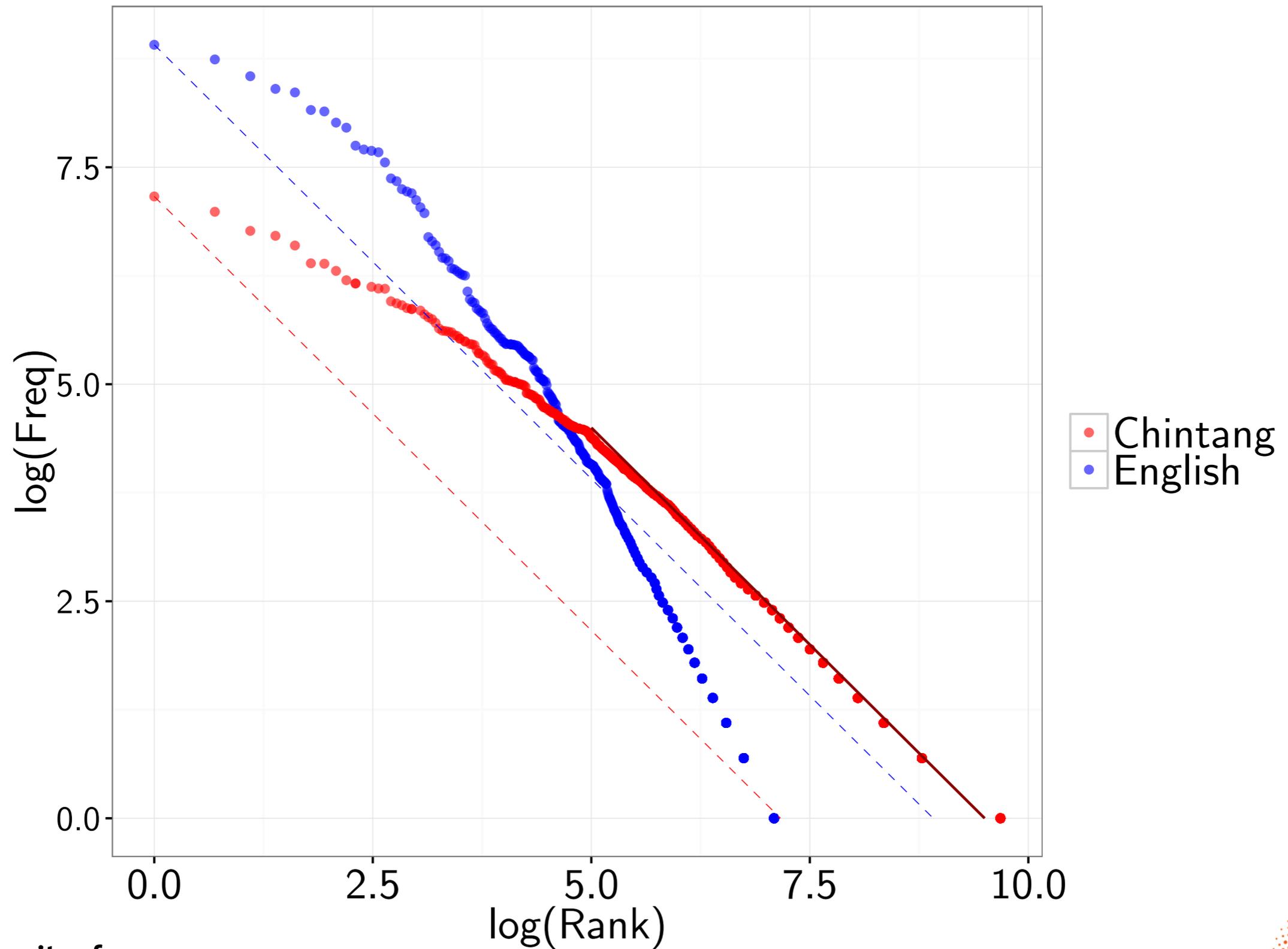
Morphological complexity: English vs.



Consequences of complexity for input distributions

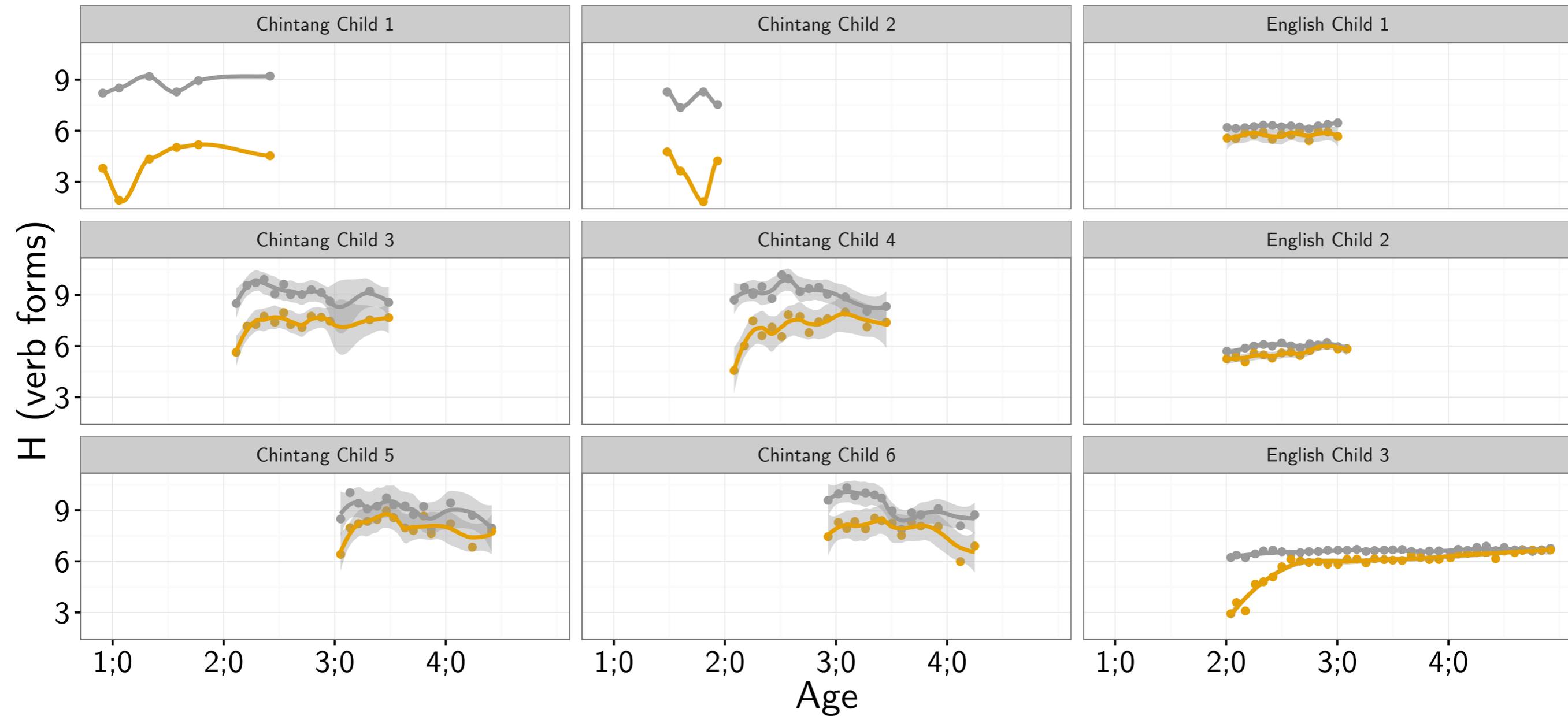


Consequences of complexity for input distributions



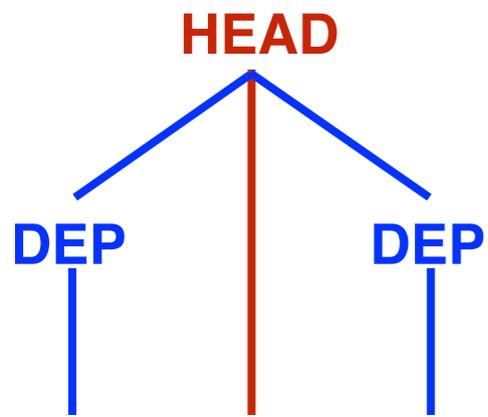
but they learn it at the same time:

surrounding speakers target children



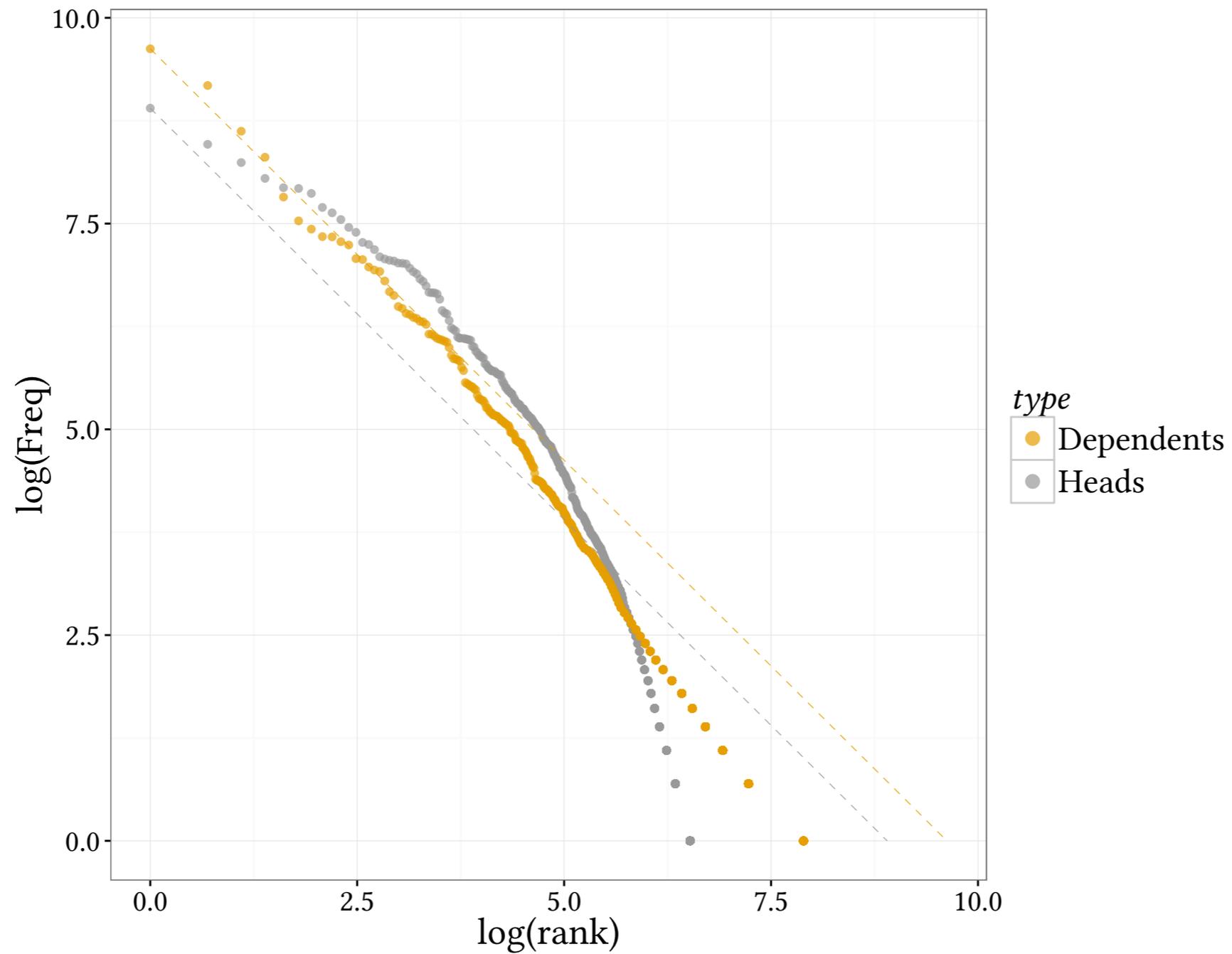
How can such a complex system be learned?

- **possible solution:** general statistical learning mechanisms
- **fact:** verbs have different elements: heads and dependents

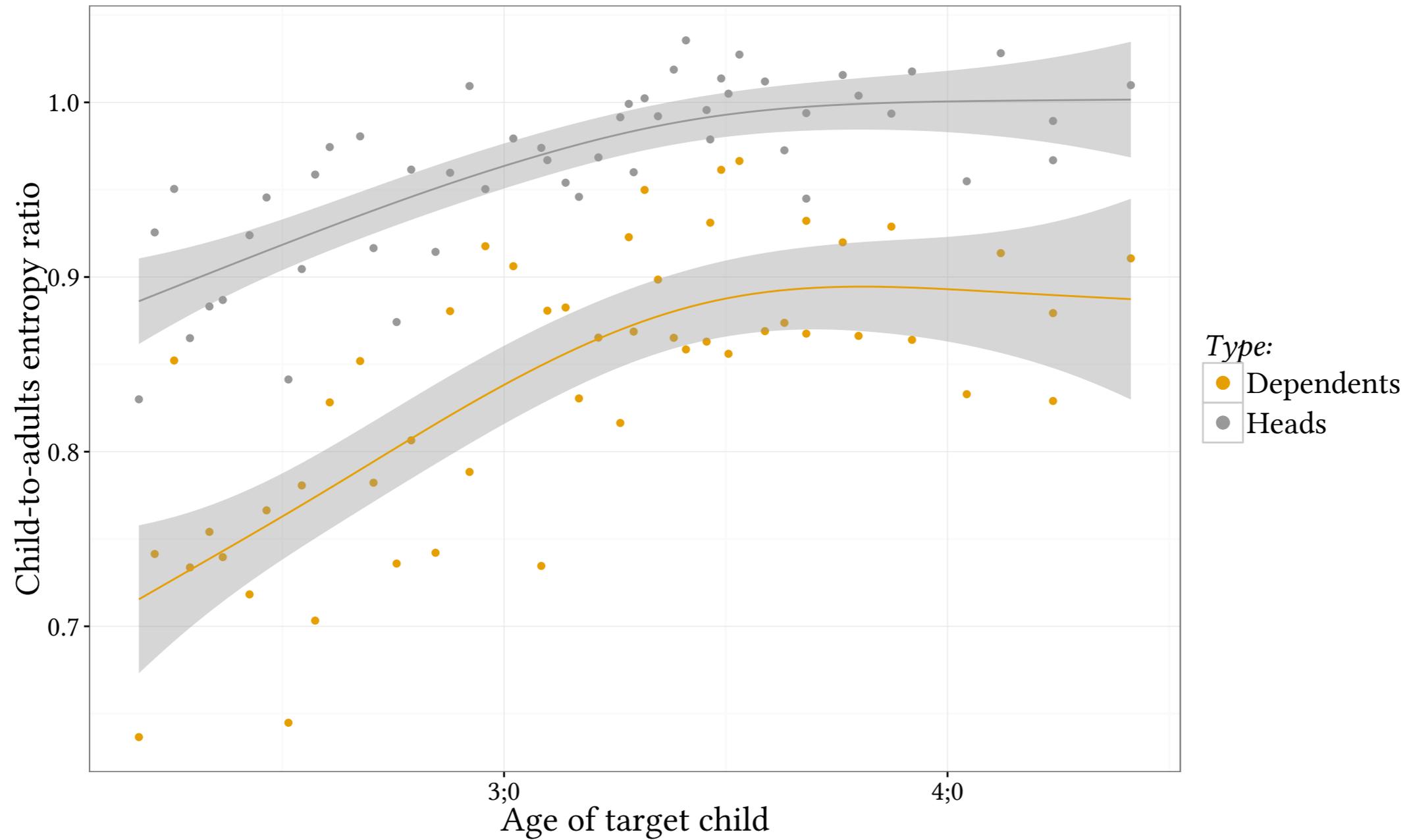


										independent stress	
										onset requirement, prosodic subcat.	
										voicing	
<i>u</i>	<i>ca</i>	<i>ŋa</i>	<i>ta</i>	<i>hai?</i>		<i>ya</i>	<i>ʔã</i>	<i>na</i>	<i>kina</i>		
3sA	eat	1sO	FOC	move.away.TR	1sO	IND.NPST	INSIST	SEQ			
[:V]	[:]	[:V]	[:X]	[:V _{2σ}]		[:V]	[:V]	[:VP]	[:XP]		
										insertion and displacement potential	
										cross-slot dependencies	
										fixed ordering	

Frequencies of heads and dependents in Chintang verbs



GAM dependents vs. heads



Conclusions

- despite extreme differences in acquisition challenges, children learn Chintang and English morphology around the same age.
- children use statistical learning strategies to learn the system.
- they adapt to the affordances of their language.
- Chintang children first learn elements with higher repetition rates within verbs, the heads, i.e. the elements in verbs that control other elements, i.e. it's dependents. Dependent elements in verbs are learned at a slower pace, there are more of them and the individual instances are less frequent. This reduces the cognitive load and makes the system learnable.



The Chintang team and funding sources

Banjade, Goma
Bibiko, Hans-Jörg
Bickel, Balthasar
Berthold, Falko
Cannedo, Laura
Domberg, Andreas
Erlach, Michael
Ghimire, Sandesh
Gaenzle, Martin
Hegetschweiler, Shirin
Harbodt, Silke
Kantak, Nicole
Klein, Felix
Krämer, Tatjana
Labs, Kristina
Lieven, Elena
Lorenz, Nicole
Paudyal, Netra
Pettigrew, Judith
Polkau, Claudia
Poppitz, Sindy
Prokopchuk, Joel
Rai, Chandra
Rai, Anita

Rai, Daya
Rai, Durga B.
Rai, Durga K.
Rai, Ganesh
Rai, Janaki
Rai, Lash K.
Rai, Rikhi
Rai, Novel K.
Rai, Shanti M.
Rai, Manoj
Rai, Rabindra
Sanjuan, Marcel
Sauppe, Sebastian
Schackow, Diana
Schikowski, Robert
Schneider Sarah
Seeg, Jenny
Strassberg, Joel
Stoll, Sabine
Vissiennon, Kodjo
Washington, Kerstin
Weymuth, Rachel
Wienholz, Anna
Wohlgemuth, Jan

Wolters, Kristin
Wunder, Saskia
Zakharko, Taras
Zeiske, Mathias



MAX-PLANCK-GESELLSCHAFT



FONDS NATIONAL SUISSE
SCHWEIZERISCHER NATIONALFONDS
FONDO NAZIONALE SVIZZERO
SWISS NATIONAL SCIENCE FOUNDATION



University of
Zurich^{UZH}