Zero argument indexing on verbs: Patterns and explanations

Laura Becker

University of Freiburg

Linguistisches Kolloquium, JGU Mainz

1. **Zero argument indexing**
   - Introduction
   - Data
   - Results
   - Development scenarios
   - 3 vs. 1, 2
   - 2(SG) imperative
   - Conclusion

2. **Outlook: Coding efficiency in DoReCo**

Zero argument indexing

In a nutshell

Zero indexes on verbs:

- Result of communicative efficiency or by-product of other diachronic processes?

	set B marker
1SG	-oñ
2SG	-ety
3SG	**-Ø**
1PL.IN	-oñ=la
1PL.EX	-oñ=l(oj)oñ
2PL	-ety=la
3PL	-ob/-o’

(1) a. tyi k-pi’ty-ä-y-**ety**.
PFV A1-wait-DT-EP-**B2**
‘I waited for you.’

b. tyi y-il-ä-**Ø** wiñik x-ixik.
PFV A3-see-DT-**B3** man CL-woman
‘The woman saw the man.’
Ch’ol (Vázquez Alvarez 2011: 25, 21)

In a nutshell

❓ Result of communicative efficiency or by-product of other diachronic processes?

It's not that simple.

There is evidence for both, depending on the phenomenon.

By-product (3 vs. 1, 2)

- differential non-development → zero argument indexing
  - a number of factors **favor** a **development** for some indexes
  - a number of factors **disfavor** a **development** for other indexes
  - ☞ the more factors add up, the stronger the crosslinguistic tendencies (cf. Cristofaro 2021)

Coding efficiency (2SG.IMPER vs. 2SG.IND)

- shorter form for more frequent / predictable messages
- non-development may be accounted for by high predictability (e.g. Haspelmath 2008)

Argument indexes

- bound argument indexes (affixes, clitics)
- no distinction between “anaphoric” and “grammatical” agreement

(2) nĩ: zihm-da **nə**-ba-ke
you house-ALL **S:2SG**-go-PFV
‘You went to the house.’ Kham (Watters 2004: 81)

(3) ku=**nash** ink awkú tx̲ána-ta asúm
and=**1SG** 1SG then become-FUT eel
‘then I will become an eel’ Sahaptin (Jansen 2010: 101,81,184)

(4) làmìnà **wó** hùn
Lamina **3SG** come
‘Lamina will come.’ Bullom So (Childs 2011: 139)

Types of zero argument indexes

Types of zero argument indexes (Siewierska 2010)

- **absolute**
- **paradigmatic**
- **allomorphic**

- **absolute**
- **non-absolute**
  (paradigmatic)
  (allomorphic)

Types of zero argument indexes

Absolute zero: general absence of the index

set B marker (ABS)	
1SG	-oñ
2SG	-ety
3SG	**-Ø**
1PL.IN	-oñ=la
1PL.EX	-oñ=l(oj)oñ
2PL	-ety=la
3PL	-ob/-o’

(5) a. tyi k-pi’ty-ä-y-**ety**.
PFV A1-wait-DT-EP-**B2**
‘I waited for you.’

b. tyi y-il-ä-**Ø** wiñik x-ixik.
PFV A3-see-DT-**B3** man CL-woman
‘The woman saw the man.’

Ch’ol (Vázquez Alvarez 2011: 25, 21)

Types of zero argument indexes

Paradigmatic zero: the index is absent in combination with certain values of another category

(6) Ngawa jayi-**ngga**=**yi**
water give=**IMP**=**P:1SG**
'Give me some water'

(7) [...] nyangula-la ga-nggu=**yi**n=nga ngayiny-jirri ngurra-nggurra
[...] sometime-LOC take-POT=**A:2SG**.**P:1SG**=DUB 1SG.DAT-ALL camp-ALL
lurrbu
return
'[...] you'll take me back to my camp sometime'

Bilinarra (Meakins and Nordlinger 2013: 302,437,436)

Types of zero argument indexes

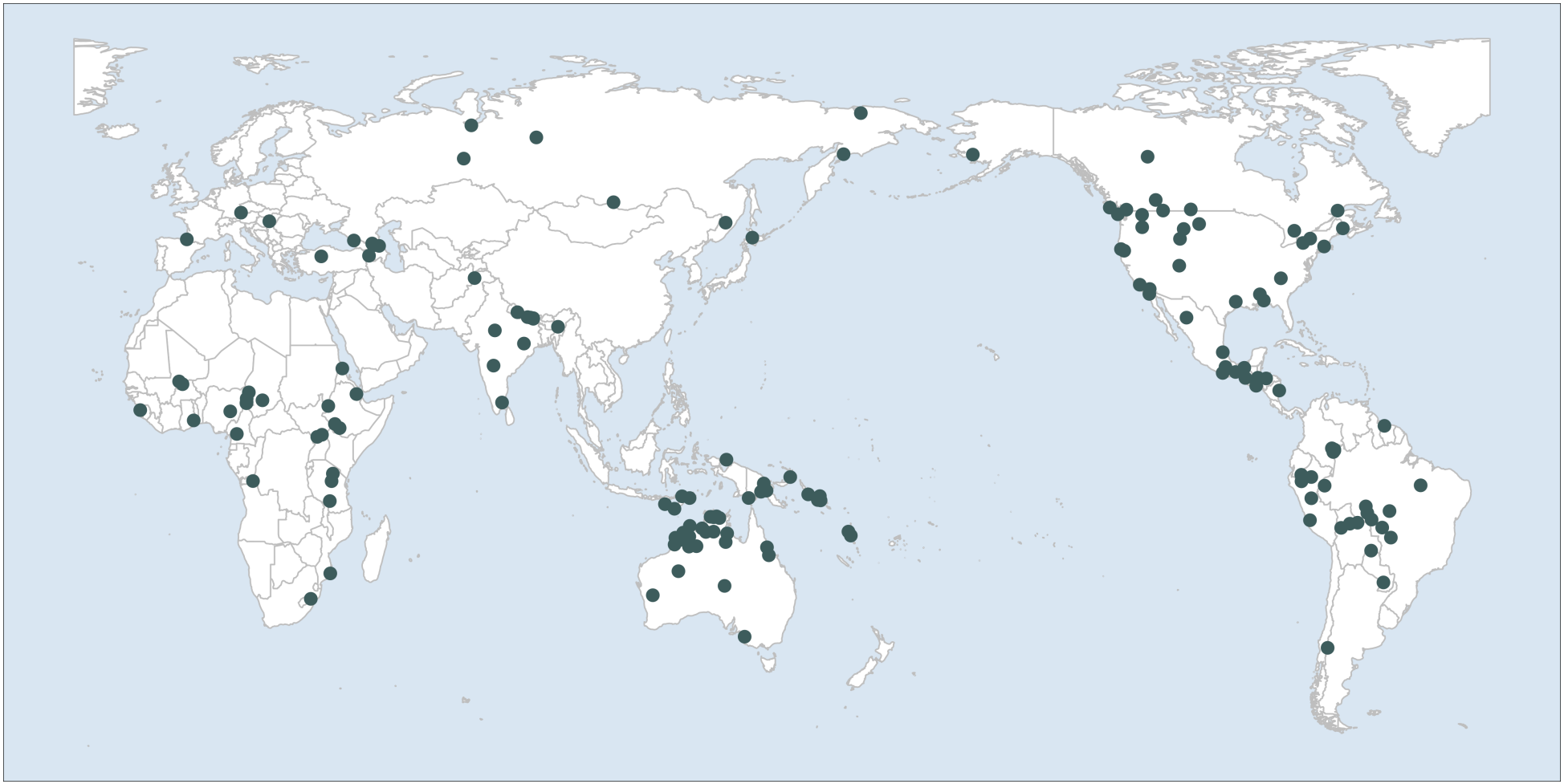
Allomorphic zero: the index is absent in certain inflection classes

(8) a. **ǐ**-wæt ‘he hit’
b. **ǐ**-jjæš ‘he entered’

(9) a. **Ø**-ùjəj ‘he went far away’
b. **Ø**-àwən ‘he goes up’ Tamashek (Heath 2011: 436)

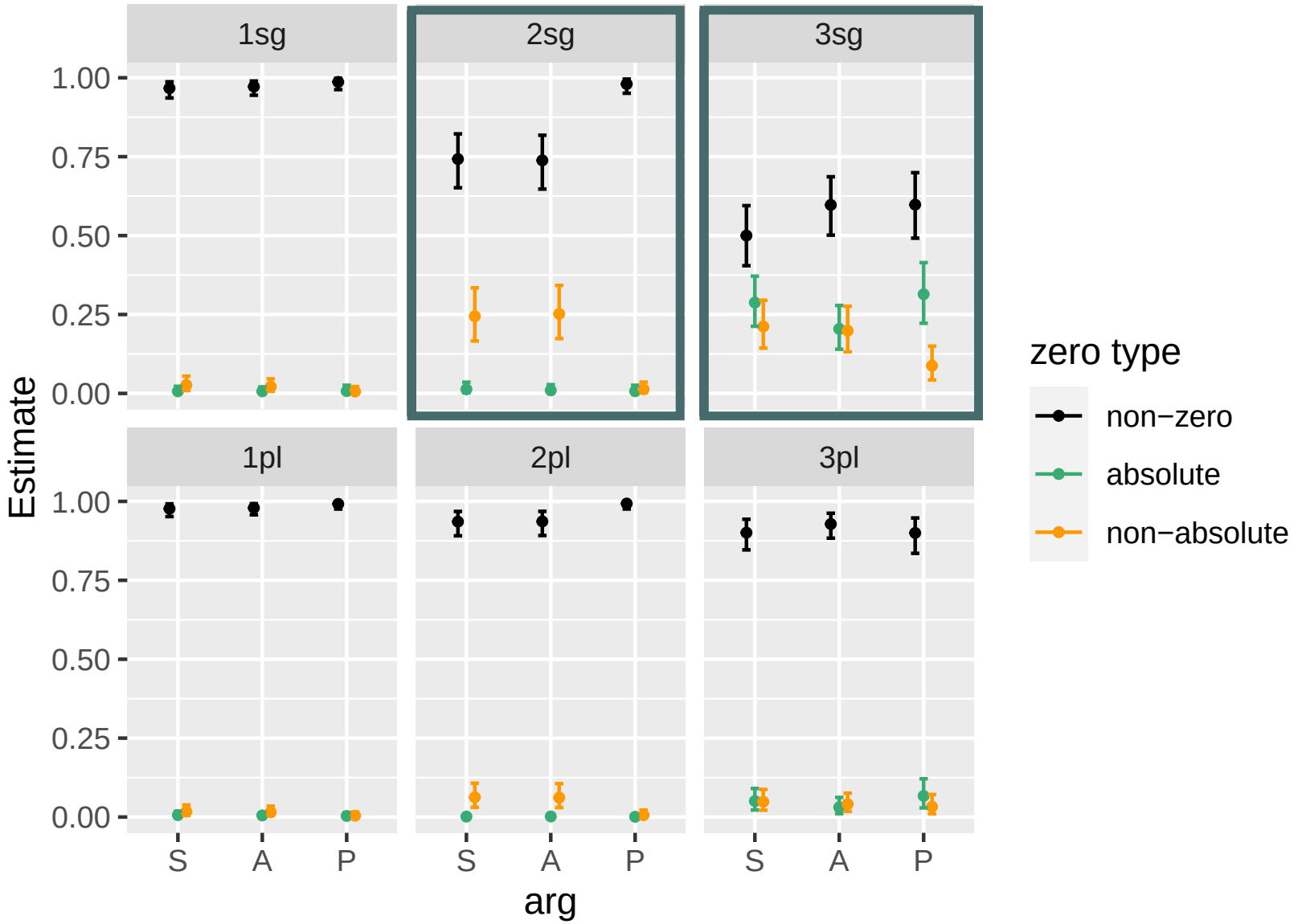
Sample

156 (out of 200) languages with argument indexes



Results:
Crosslinguistic distributions

Modelling zero forms across person and arguments



Bayesian regression with phylogenetic and areal controls Guzmán Naranjo and Becker (cf. 2021)

Explaining the patterns:
How do zero forms develop?

Development of zero forms

- differential phonetic reduction (✔)
  not likely driven by communicative efficiency (cf. Seržant 2021)

- differential structural reduction (✖)

☞ reanalysis is probably what happens instead of structural reduction (✔)

- **differential non-development** ✔

Reanalysis → differential loss

Watkins's law

- a 3rd person index is reanalysed as part of the stem

☞ 3rd person zero index (Watkins 1962; Koch 1995)

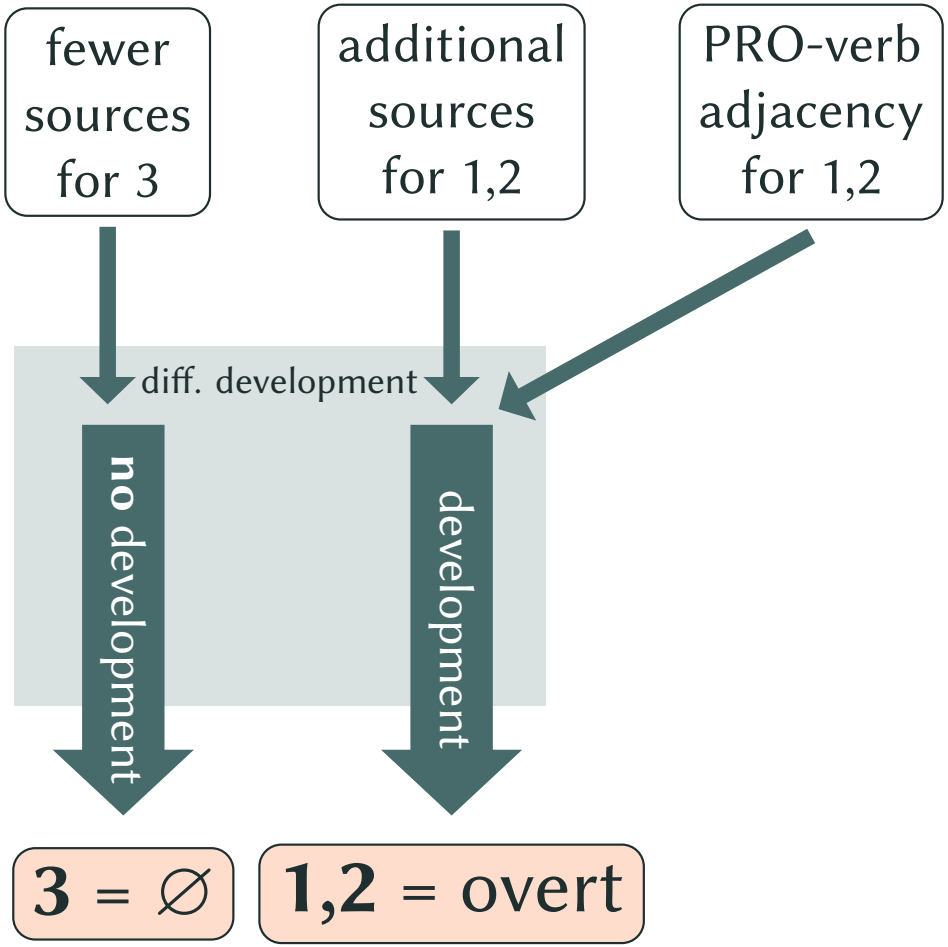
	Proto-Iranian						Persian		
1SG	*as-**mi**	>	*as-**m**			>	*ast-**m**	>	hast-**am**
2SG	*as-**i**	>	*as-**i**			>	*ast-**i**	>	hast-**i**
3SG	*as-**ti**	>	*as-**t**	>	*ast-∅	>	*ast-∅	>	hast-∅

(Watkins 1962: 94)

3 vs. 1, 2

Non-development resulting from various factors

3 vs. 1, 2


fewer sources for 3
additional sources for 1,2
PRO-verb adjacency for 1,2
diff. development
no development
development
3 = ∅
1,2 = overt

Fewer sources for 3 → differential development

- languages may not have third person pronouns

☞ only 1st and 2nd person pronouns developed into bound person indexes in Tabasaran (Lezgic) (Helmbrecht 1996; Bogomolova 2018)

	NOM	ERG	DAT
1SG	**uzu**		**uzuz**
2SG	**uvu**		**uvuz**
3SG	dumu	duʁu	duʁaz

	A	P	EXP/REC
1SG	-za	**-zu**	**-zuz**
2SG	-va	**-vu**	**-vuz**
3SG	-Ø	-Ø	-Ø

Additional sources for 1, 2 → differential development

cislocative markers > 1, 2 object indexes

(cf. Konnerth and Sansò 2021; Cristofaro 2021)

- a cislocative marker ('hither') develops into an object marker, but only for speech act participants

(10) [...] e-nūt a-ka-prék a-monít abàng=ke saikél
[...] one-CLF:HUM.SG POSS-NMLZ-be.different POSS-man NPDL=TOP bicycle
nang=ardòn-si vàng-lò ...
CIS=ride-NF:RL come-RL ...
'[...] another person riding on a bicycle came, ...' Karbi (Konnerth 2015: 35)

(11) [...] **nang**=ke-che-arjū-lò pēi=pen pō
[...] **P:1/2**=IPFV-RR-ask-RL mother=with father
'[...] we are asking you, mother and father' Karbi (Konnerth 2015: 31)

Paradigm formation ≠ a single process

3 vs. 1,2: Different sociopragmatic pressures

- more sociopragmatic pressures on 1,2 vs. 3 forms (Konnerth 2015: 4):
  - avoidance
  - special highlighting for seeking empathy

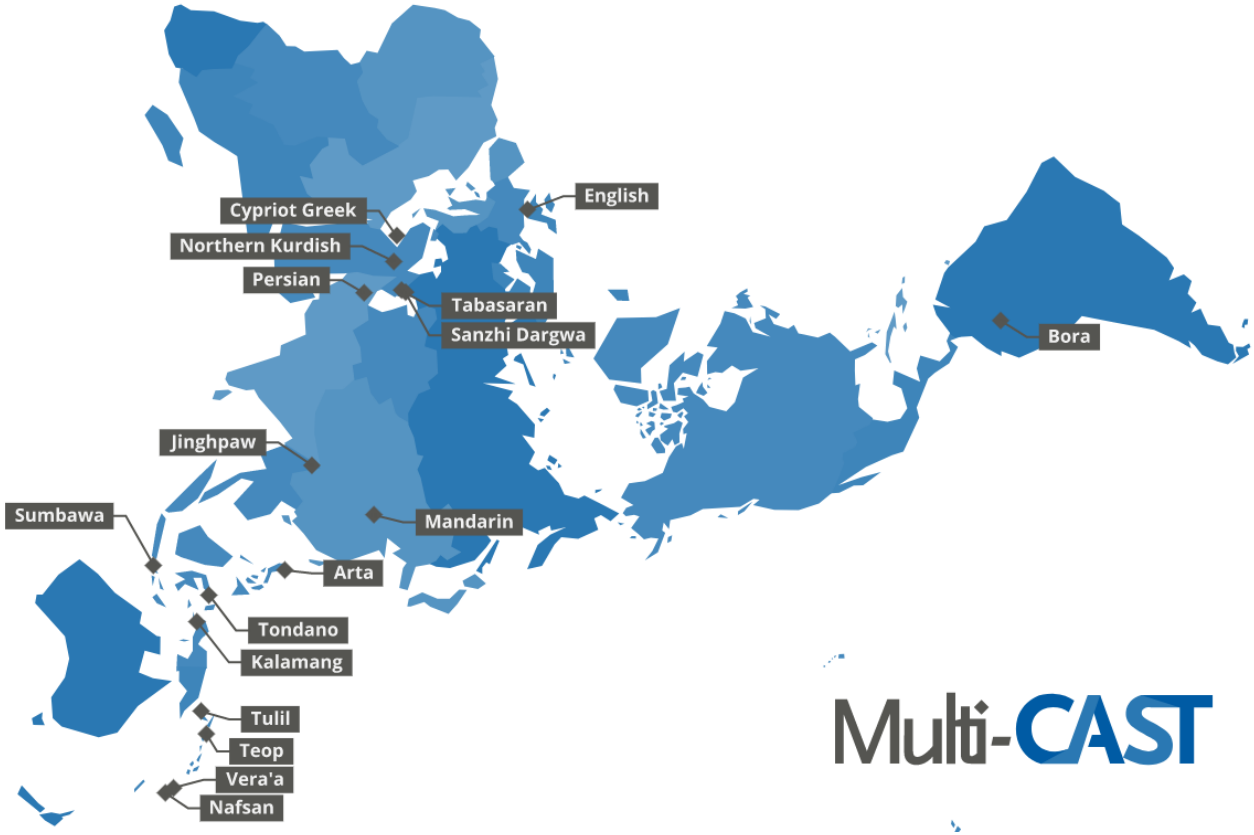
3 vs. 1,2: Independent developments

- 1st, 2nd and 3rd person indexes can develop independently (Mithun 1991)

Plains Cree		
1SG	**ni**pimipahtan	'I run'
2SG	**ki**pimipahtan	'you run'
3SG	pimipahta**w**	'he/she runs'
1PL.EX	**ni**pimipahtanan	'we run'
1PL.IN	**ki**pimipahtananaw	'we run'
2PL	**ki**pimipahtanawaw	'you run'
3PL	pimipahta**wak**	'they run'

Multilingual Corpus of Annotated Spoken Texts (Haig and Schnell 2021)

- spontaneous spoken data (mostly monologic) from 17 languages (27 500 clause units)
- ! keeps track of discourse referents including pro drop



PRO-verb adjacency for 1, 2 → differential development

- MultiCAST: probability of PRO is lower for 3

		S		A		P	
1	pro	1172	**(0.69)**	1148	**(0.62)**	240	**(0.84)**
	∅	523		713		45	
2	pro	340	**(0.56)**	280	**(0.41)**	108	**(0.95)**
	∅	271		402		6	
3	pro	2071	**(0.25)**	1150	**(0.24)**	785	**(0.10)**
	∅	3145		2798		1347	
	lex	2965		927		5856	

PRO-verb adjacency for 1, 2 → differential development

MultiCAST

- the probability for 1, 2 PRO to occur adjacent to verbs is higher than for 3 PRO
- especially for P arguments

☞ this may facilitate the development of indexes for 1, 2

	verb – PRO					
	S		A		P	
	N	prob	N	prob	N	prob
1	38	0.02	24	0.01	**100**	**0.35**
2	6	0.01	20	0.03	**92**	**0.81**
3	30	0.004	12	0.002	**713**	**0.09**

PRO-verb adjacency for 1, 2 → differential development

MultiCAST

- the probability for 1, 2 PRO to occur adjacent to verbs is higher than for 3 PRO
- smaller difference for A and S arguments

	PRO – verb					
	S		A		P	
	N	prob	N	prob	N	prob
1	**938**	**0.55**	**917**	**0.49**	43	0.15
2	**212**	**0.35**	**181**	**0.27**	19	0.17
3	**1777**	**0.21**	**1020**	**0.21**	84	0.01

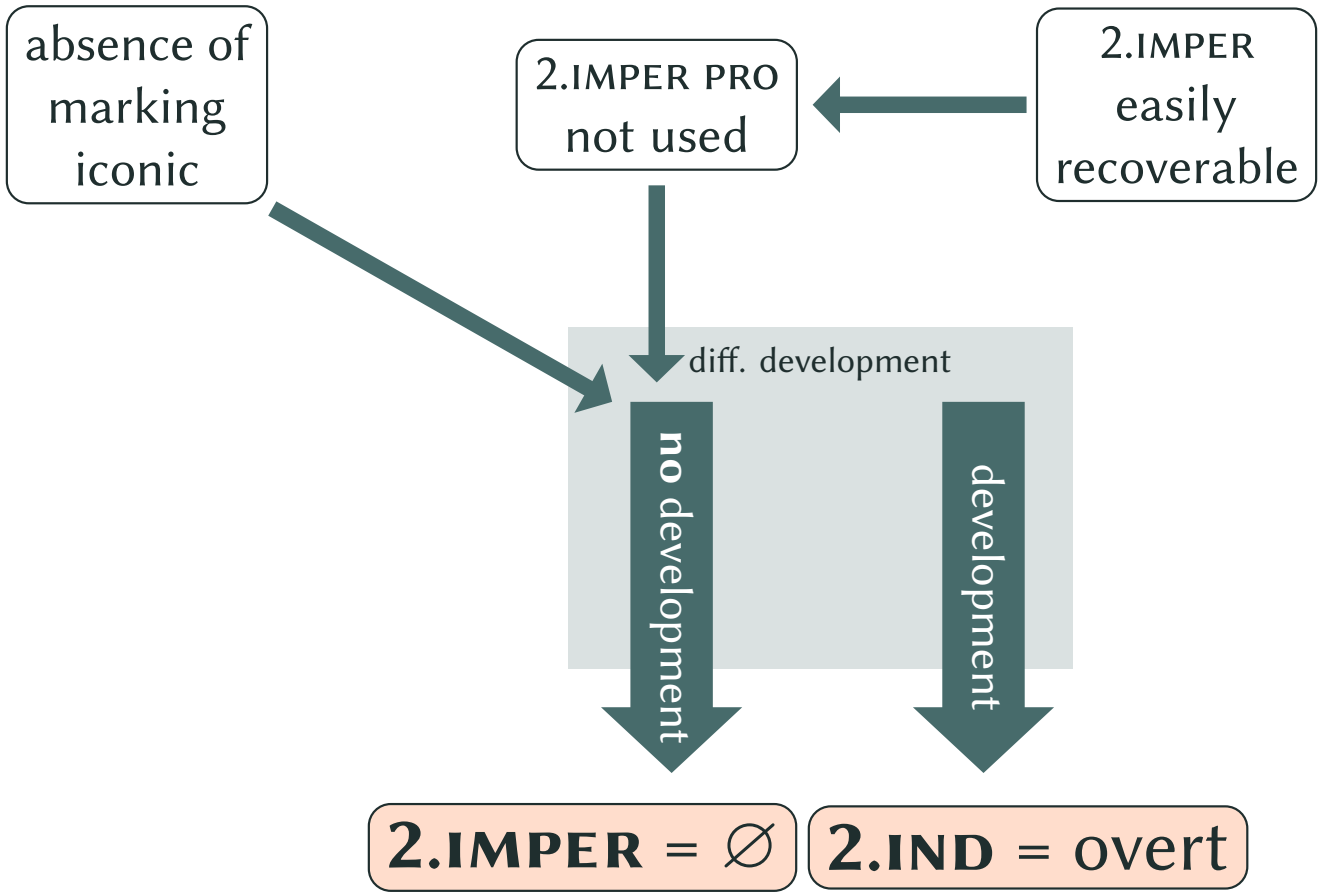
2(SG) imperative

vs.

2(SG) indicative

Non-development (partially) driven by communicative efficiency?

2(SG) imperative vs. 2(SG) indicative



Are imperatives associated with zero forms?

- zero forms are likely to be used for IMPER.2SG (29% of the sample)
- but IMPER.2SG as such does not appear to be strongly associated with zero forms

- confirms previous findings (Aikhenvald 2010; van der Auwera et al. 2013)
  - Aikhenvald (2010: 18): zero marking in **one third** of the languages of the world
  - Bybee, Perkins, and Pagliuca (1994): 11 out of 136 languages (**8%**) showed zero-marked IMPER.2SG
  - van der Auwera et al. (2013): 122 out of 548 languages (**22%**) have no morphologically dedicated 2nd person imperative forms

No suitable source for 2sg.imper → non-development

- the subject is necessarily the addressee and most recoverable from the discourse situation
  (Sadock and Zwicky 1985; Levshina 2018; Aikhenvald 2010; Nikolaeva 2007)

☞ in most languages, the subject of IMPER.2SG does not have to be expressed
(Aikhenvald 2010: 92, Sadock and Zwicky 1985: 171)

MultiCAST

☞ imperative forms are indeed much more likely to occur with no 2nd person pronoun in the same clause

	overt PRO	no PRO
SG	37	247 (**0.87**)
PL	10	29 (**0.74**)

Iconicity → morphological reduction?

- **shortest** possible form → **directness**, **urgency** (Aikhenvald 2010: 46)

(12) a. Eh dyilbi, banyi **urldi**-**ga**!
eh old.man, here **come**-**IMPER**
'Come over here old fellow!' (Hercus 1999: 55)

b. Banhi **urldi**!
here **come**
'Come here!' Wirangu (Hercus 1999: 117)

(13) **diga**-**(ja)**
'**eat**-**(IMPER)**!'
'eat!' Udihe (Nikolaeva and Tolskaya 2011: 221-222)

(14) ʔam **nɔ́h**!
2SG fall.**APPR**
'(Watch out,) you'll fall!' Hup (Epps 2008: 631)

Conclusion

Multiple factors

- many different factors contribute to the distribution of zero argument indexes
- the more factors add up and do not cancel each other out, the stronger the crosslingusitic trend

☞ communicative coding efficiency alone can probably not account for crosslinguistic preferences

Differential non-development > zero

- non-development scenario most relevant for the development of (absolute) zero argument indexes
  - ☞ 3rd person > 1st, 2nd person
  - ☞ P > S, A
  - ☞ 2SG.IMP > 2SG.IND

Outlook

Coding efficiency in person-number indexes

DoReCo

- corpus collection with data from 51 languages
- spoken language
- time-aligned (on the phonological level)
- morphological annotation
- **today**: 4 languages (Beja, Baïnounk Gubëeher, Dolgan, Texistepec Popoluca)

`https://doreco.huma-num.fr/` (Seifart, Paschen, and Stave 2022)

Annotation

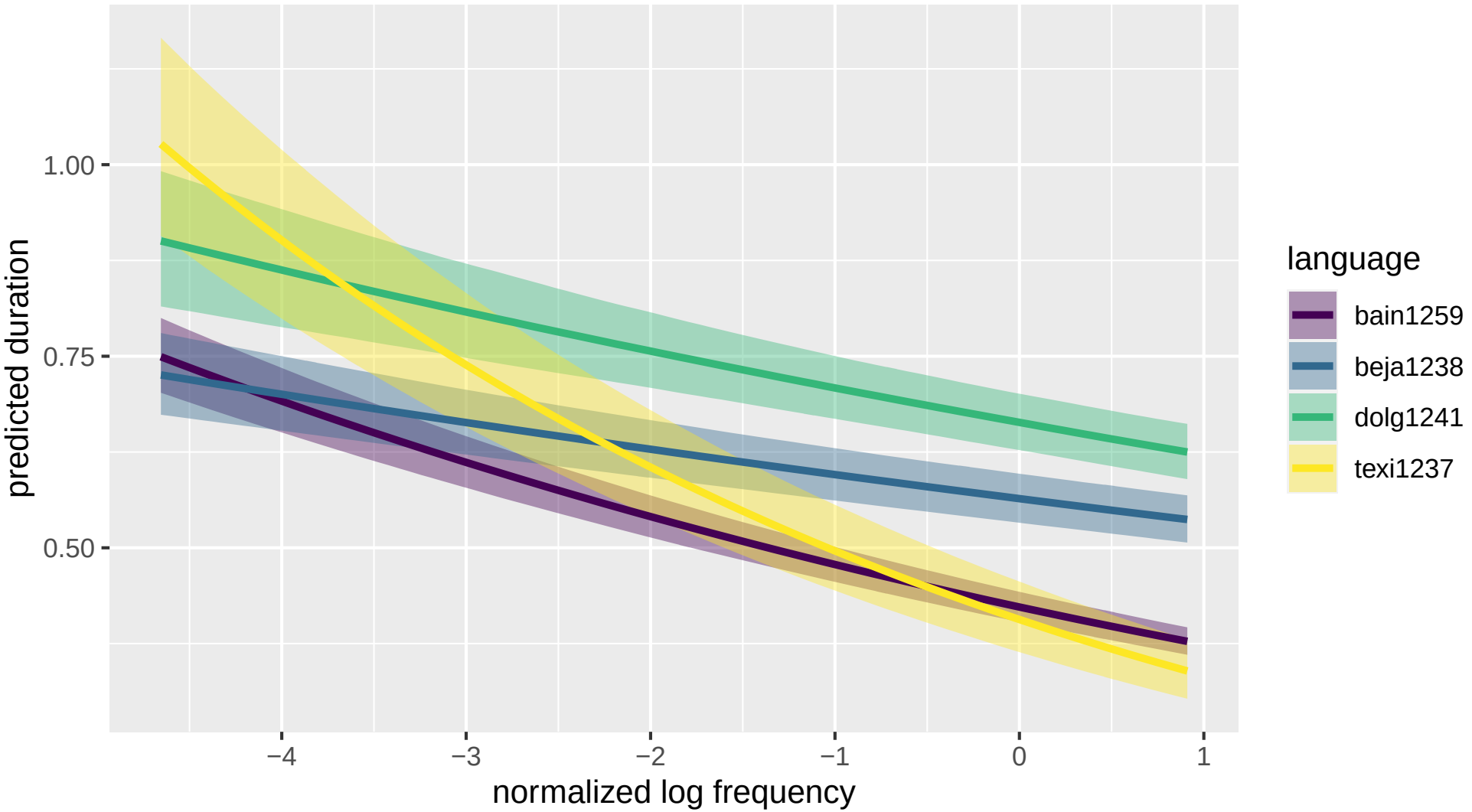
Beja: 1SG

(15) ani i=nafs=ijoː=d **hadiːd-ani**
1SG.NOM DEF.M=SOUL=POSS.1SG.GEN=DIR **TALK-IPFV.1SG**
'I am talking about myself'
(0035_DoReCo_doreco_beja1238_BEJ_MV_CONV_01_RICH)
duration: **0.668 s**
speech rate: 2 phonemes / 0.668 s = **2.99 phonemes/s**
frequency: **341**

(16) naː=t hoː **eː-tgiːm**
thing=INDEF.F LOC **1SG-ignore.MID.IPFV**
'except that I know nothing about that.'
(0060_DoReCo_doreco_beja1238_BEJ_MV_CONV_01_RICH)
duration: **0.500 s**
speech rate: 2 phonemes / 0.500 s = **4.00 phonemes/s**
frequency: **341**

The effect of frequency on duration

Preliminary results



Thank you!

References I

Aikhenvald, Alexandra (2010). *Imperatives and Commands.* Oxford: Oxford University Press.

Bogomolova, Natalia (2018). "The Rise of Person Agreement in East Lezgic: Assessing the Role of Frequency." In: *Linguistics* 56.4, pp. 819–844.

Bybee, Joan, Revere Perkins, and William Pagliuca (1994). *The Evolution of Grammar. Tense, Aspect, and Modality in the Languages of the World.* Chicago: The University of Chicago Press.

Childs, George Tucker (2011). *A Grammar of Mani.* Mouton Grammar Library 54. Berlin: De Gruyter.

Cristofaro, Sonia (2021). "Typological Explanations in Synchrony and Diachrony: On the Origins of Third Person Zeroes in Bound Person Paradigms." In: *Folia Linguistica* 55.s42-s1, pp. 25–48.

Epps, Patience (2008). *A Grammar of Hup.* De Gruyter.

Guzmán Naranjo, Matías and Laura Becker (2021). "Statistical Bias Control in Typology." In: *Linguistic Typology* 26.3, pp. 605–670.

Haig, Geoffrey and Stefan Schnell (2021). "Multi-CAST: Multilingual Corpus of Annotated Spoken Texts." Bamberg.

Haspelmath, Martin (2008). "Creating Economical Morphosyntactic Patterns in Language Change." In: *Linguistic Universals and Language Change*. Ed. by Jeff Good. Oxford: Oxford University Press, pp. 185–214.

Heath, Jeffrey (2011). *A Grammar of Tamashek (Tuareg of Mali).* Berlin: De Gruyter.

Helmbrecht, Johannes (1996). "The Syntax of Personal Agreement in East Caucasian Languages." In: *STUF - Language Typology and Universals* 49.2, pp. 127–148.

Hercus, Luise A. (1999). *A Grammar of the Wirangu Language from the West Coast of South Australia.* Pacific Linguistics.

Jansen, Joana Worth (2010). "A Grammar of Yakima Ichishkiin/Sahaptin." PhD thesis. Eugene, OR: University of Oregon.

Koch, Harold (1995). "The Creation of Morphological Zeros." In: *Yearbook of Morphology 1994*. Ed. by Geert Booij and Jaap van Marle. Dordrecht: Springer, pp. 31–731.

Konnerth, Linda (2015). "A New Type of Convergence at the Deictic Center: Second Person and Cislocative in Karbi (Tibeto-Burman)." In: *Studies in Language. International Journal sponsored by the Foundation "Foundations of Language"* 39.1, pp. 24–45.

Konnerth, Linda and Andrea Sansò (2021). "Towards a Diachronic Typology of Individual Person Markers." In: *Folia Linguistica* 55.s42-s1, pp. 1–24.

References II

Levshina, Natalia (2018). "Towards a Theory of Communicative Efficiency in Human Languages." Habilitationsschrift. Leipzig: Leipzig University.

Meakins, Felicity and Rachel Nordlinger (2013). *A Grammar of Bilinarra, an Australian Aboriginal Language of the Northern Territory*. Berlin: De Gruyter.

Mithun, Marianne (1991). "The Development of Bound Pronominal Paradigms." In: *Language Typology 1988: Typological Models in the Service of Reconstruction*. Ed. by Winfred Lehmann and Helen-Jo Jakusz Hewitt. Amsterdam: Benjamins, pp. 85–104.

Nikolaeva, Irina (2007). *Finiteness: Theoretical and Empirical Foundations*. Oxford: Oxford University Press.

Nikolaeva, Irina and Maria Tolskaya (2011). *A Grammar of Udihe*. De Gruyter.

Sadock, Jerrold and Arnold Zwicky (1985). "Speech Act Distinctions in Grammar." In: *Language Typology and Syntactic Description. Volume 1*. Ed. by Timothy Shopen. Cambridge: Cambridge University Press, pp. 155–196.

Seifart, Frank, Ludger Paschen, and Matthew Stave, eds. (2022). *Language Documentation Reference Corpus (DoReCo) 1.2*. Berlin & Lyon: Leibniz-Zentrum Allgemeine Sprachwissenschaft & laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2).

Seržant, Ilja (2021). "Cyclic Changes in Verbal Person-Number Indexes Are Unlikely." In: *Folia Linguistica* 55.s42-s1, pp. 49–86.

Siewierska, Anna (2010). "Person Asymmetries in Zero Expression and Grammatical Functions." In: *Essais de Linguistique Generale et de Typologie Linguistique Offerts Au Professeur Denis Creissels à l'occasion de Ses 65 Ans*. Ed. by Franck Floricic. Paris: Presses de l'École Normale Supérieure, pp. 425–438.

van der Auwera, Johan et al. (2013). "The Morphological Imperative." In: *The World Atlas of Language Structures Online*. Ed. by Matthew S. Dryer and Martin Haspelmath. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Vázquez Alvarez, Juan Jesús (2011). "A Grammar of Chol, a Mayan Language." PhD thesis. Austin, TX: University of Texas.

Watkins, Calvert (1962). *Indo-European Origins of the Celtic Verb*. Dublin: Dublin Institute for Advanced Studies.

Watters, David (2004). *A Grammar of Kham*. Cambridge: Cambridge University Press.