



HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Morphological complexity meets sociolinguistics: A typological approach

Structural Complexity in Natural Language(s) (SCNL)

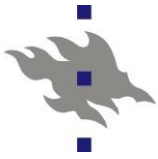
International Workshop

Paris, 30-31 May 2016

Kaius Sinnemäki,

Helsinki Collegium for Advanced Studies

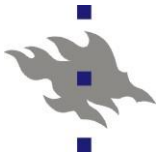
University of Helsinki (kaius.sinnemaki@helsinki.fi)



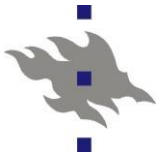
Background

- Language usage varies along several social parameters
 - Age, gender, socioeconomic status, class, ...
 - Research mostly on different varieties of one language.
 - E.g. Labov (1996) focused on English varieties in New York.



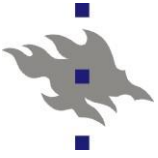


- Is there any sociohistorical reason why
 - the English say
 - *day* vs. *day-s*
 - whereas the Faroese say
 - *dagur* ['dɛavʊr] 'day' vs. *dagar* ['dɛ:ar] 'days'
- (cf. Dammel & Kürschner 2008)

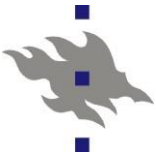


- Or why the Finns say
 - *tupa* 'house, cottage' vs. *tuva-n* 'house.gen'
 - (and not *tupa-n* 'house.gen')

- but the Estonians say
 - *tuba* 'room, chamber' vs. *toa* 'room.gen'
 - (and not *tuba-n* or *tuva-n*)?



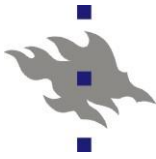
- The basic question here:
 - How much can social variation condition grammar? Does grammar adapt to the sociocultural environment?



- The mainstream view:

“There is no correlation whatsoever between phonological structure (or, for that matter, any aspect of linguistic structure) and the environment.” (Kaye 1989: 48)

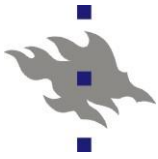
“[V]irtually all linguists today would agree that there is no hope of correlating a language's gross grammatical properties with sociocultural facts about its speakers.” (Newmeyer 2002: 361)



“If you simplify a language’s structure in one place, you are likely to complicate it somewhere else [... This works] for externally-motivated change as well as for internally-motivated change.” Sarah Thomason (2008).

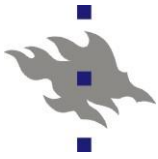
“In general ... attempts to link language structure with extralinguistic factors are almost intrinsically suspect.”

Robert Ladd et al. (2015).



- In functionalist approaches, grammar adapts
 - to preferences in language use (e.g., Bybee 2010).
 - Sociocultural change affects patterns of usage
 - and ultimately to cultural ecology in language acquisition, diachrony, and evolution (cf. Givón 2009).
 - E.g. societies of intimates vs. strangers.

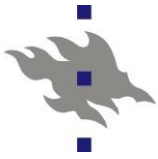
- Increasing interest from a cross-linguistic perspective:
 - See reviews by Nettle (2012) and Ladd et al. (2015).



Aim in this paper

- How could we start problematizing the question and begin to answer it?

- I discuss here one example:
 - Morphological complexity and language contact.
 - How to start connecting morphological complexity with sociolinguistics.

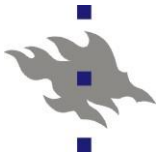


Type of complexity

- Complexity = number and type of parts and their interactions.

- Here focus:
 - complexity as the number of parts in a system.

 - Many typological studies focus on inventories; very few survey interactions or transparency (see Kusters 2003):
 - transparency in morphology (e.g. fusion) may vary depending on the degree of language contacts by adults.
 - More second language (L2) learners → less fusion.



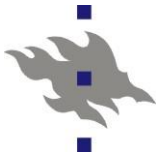
Language contact and complexity

- Trudgill (2011): Variation in sociolinguistic situation correlates with the complexity of linguistic structure.

Large adult L2 pop. (“high contact”)	→	Simplification
Small adult L2 pop. (“low contact”)	→	Maintenance or development of complexity
Child multilingualism	→	Complexification

- Rationale:

- Imperfect learning by adult second language learners.
- Morphology in particular is difficult for L2 learners (e.g. Parodi et al. 2004).

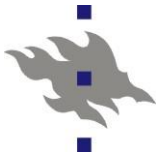


Inflectional synthesis - hypothesis

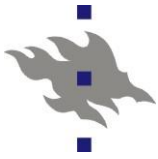
- Kusters (2003):
 - Language varieties more affected by massive L2 by adults prefer fewer morphological categories on the verb.

- Trudgill (to appear):
 - Polysynthetic language tend to be spoken by relatively small and non-industrialized tribal communities.

- DeLancey (2014):
 - Development of new morphological categories on the verb in a low-contact 'Hill culture' language Lai but not in a high-contact 'Valley culture' language Boro.



- Question:
 - Does the degree of inflectional synthesis on the verb depend on the proportion of L2 speakers in the community (following Bentz & Winter 2012)?
 - (See also Lupyan & Dale 2010).

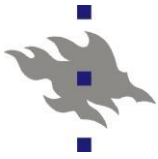


Data

- Inflectional synthesis =
 - Number of **categories per word** in maximally inflected verb forms.
 - For instance, two categories counted for English: agreement (-s) and tense (-ed).

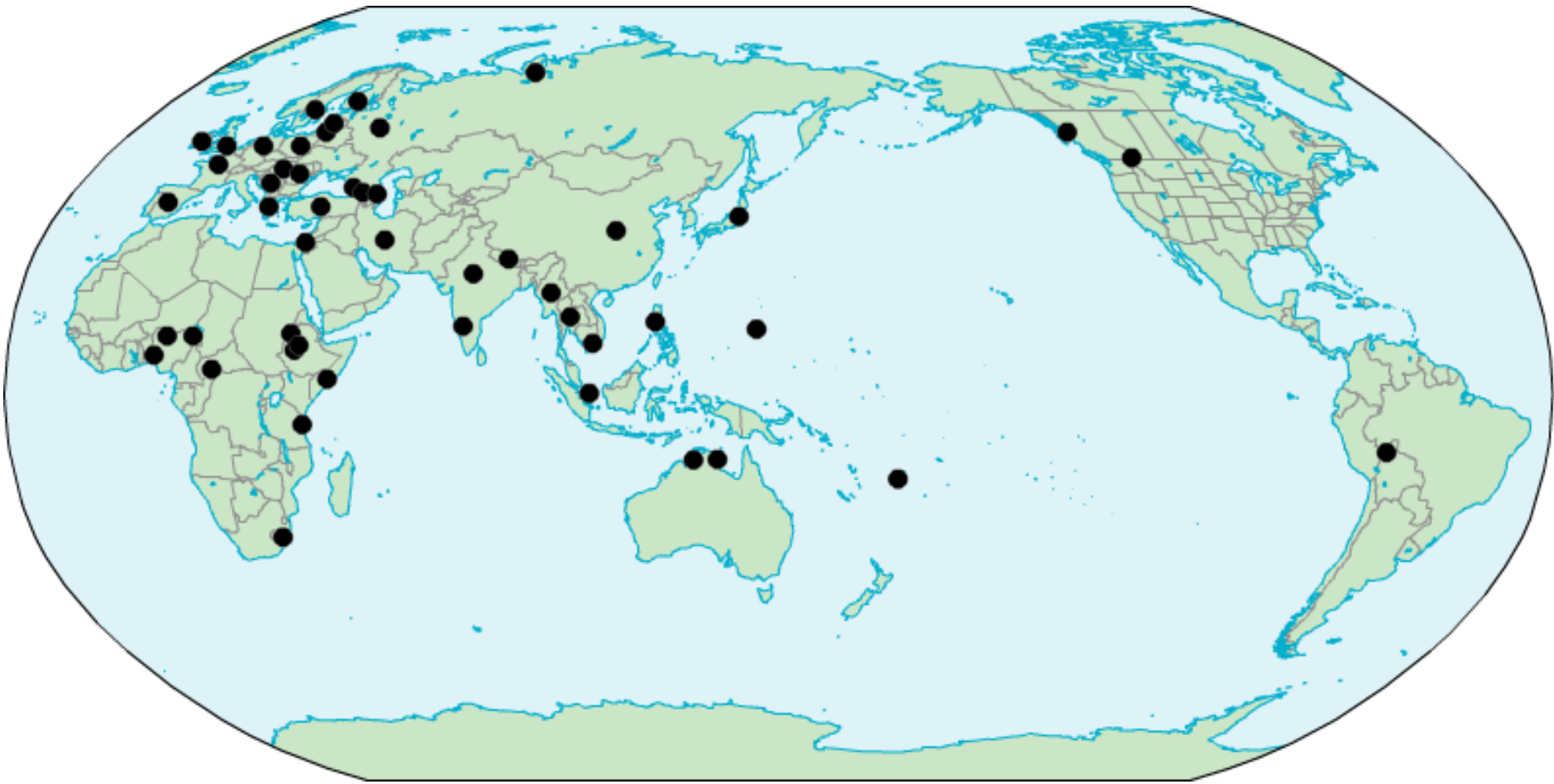
- Data from the *Autotyp* database (Nichols et al. in press).
 - Kindly provided by Balthasar Bickel, which is gratefully acknowledged.

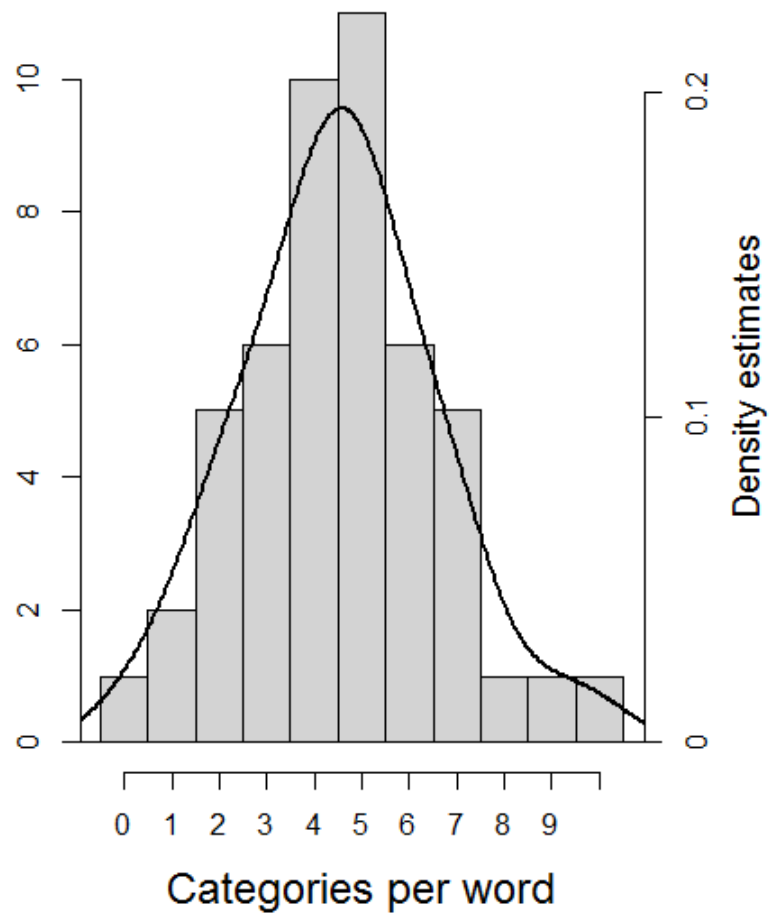
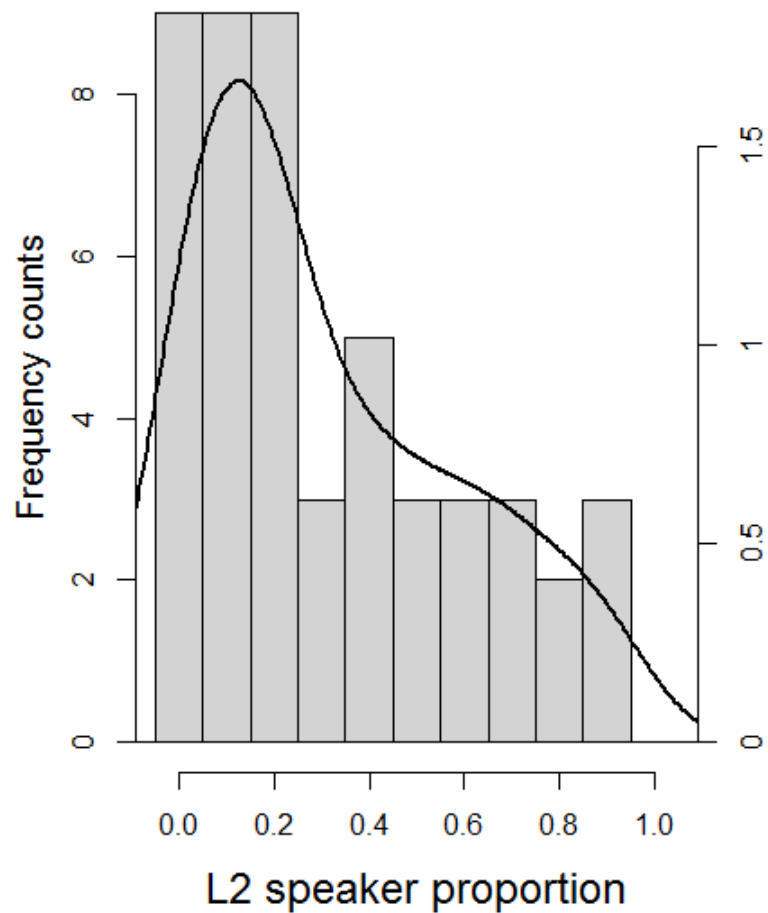
- Sample on inflectional synthesis: 268 languages.

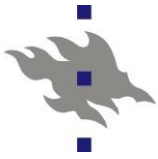


- The number of speakers was taken from the Ethnologue (Lewis et al. 2016) and Christian Bentz's database that he kindly shared.
 - Bentz's database contains information about the number of native language (L1) and second language (L2) speakers for 231 languages.

- The combined database for inflectional synthesis and the number of speakers contains data on 49 languages.
 - The sample is geographically very biased.
 - Overall: reasonably reliable data available for L1 **but not for L2.**





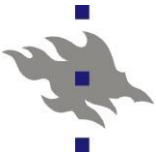


On statistical testing

- How to deal with the confounding effects of inheritance and diffusion?
 - Usually through sampling, e.g. one language per family.

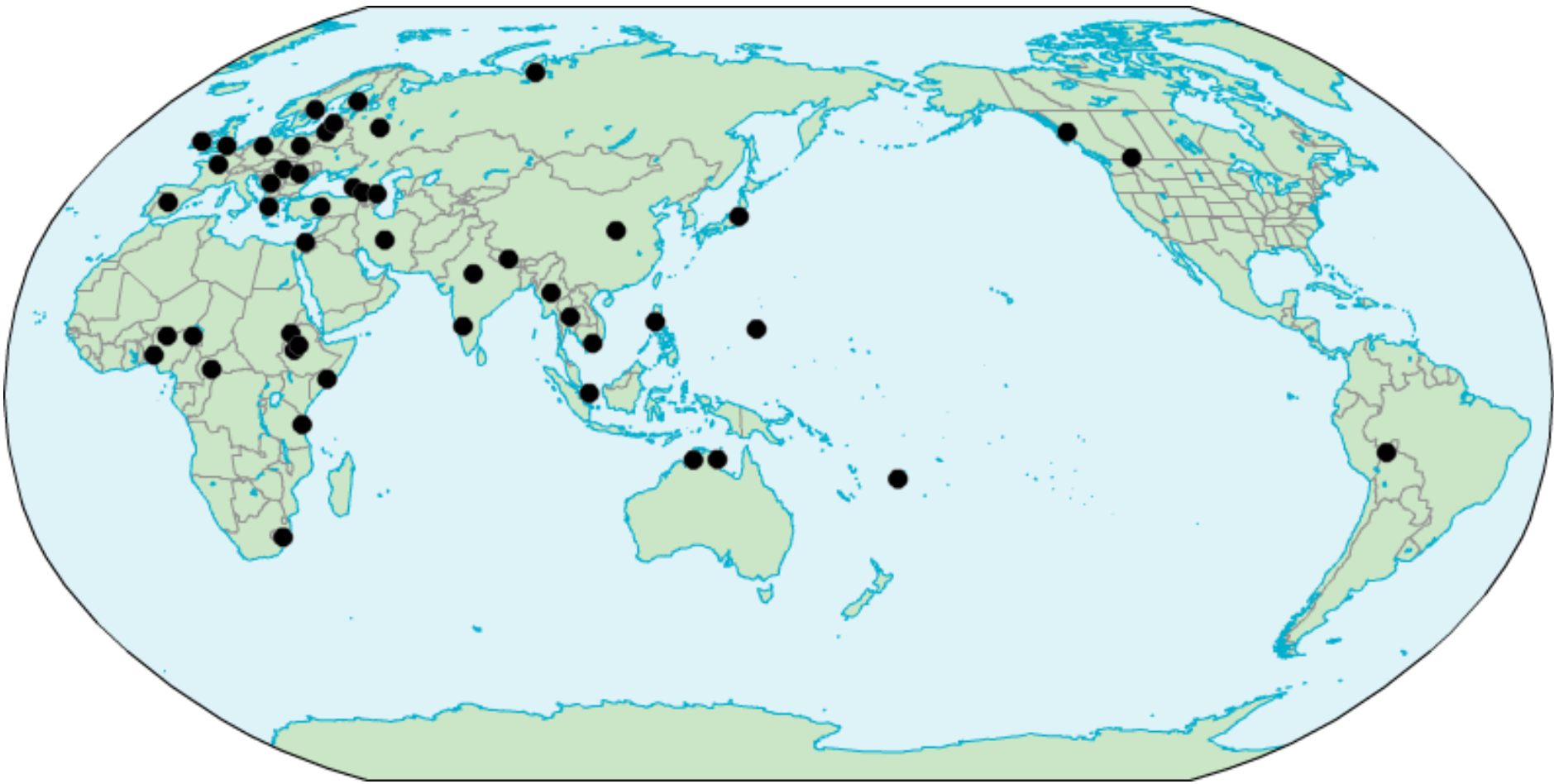
- 300+ stocks (highest level genealogical units in *Autotyp*; Nichols et al. 2013).
 - impossible to build family as a variable in multiple regression modeling (Sinnemäki 2010).

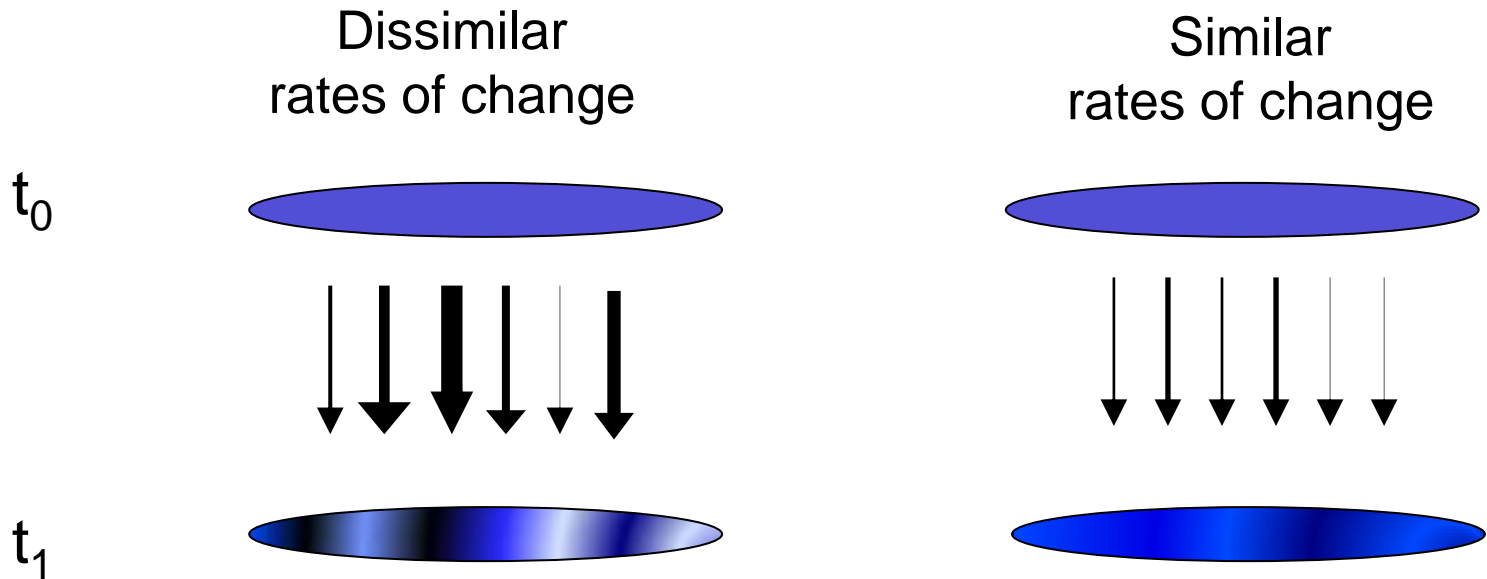
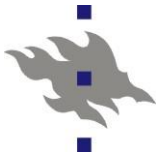
- I use linear mixed effects modeling to test the hypothesis (see Jaeger et al. 2011; Bentz & Winter 2012).
 - Enables families and areas to be built as variables in the model → their effect can be evaluated.



- Two types of random effects:
 - Random slopes =
 - The predictor's effect can vary across genealogical units and areas, thus assuming that rates of change vary between different families and between different areas.

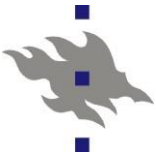
 - Random intercepts =
 - The predictor's effect is the same across families and areas, thus assuming that rates of change are similar between different families and between different areas.



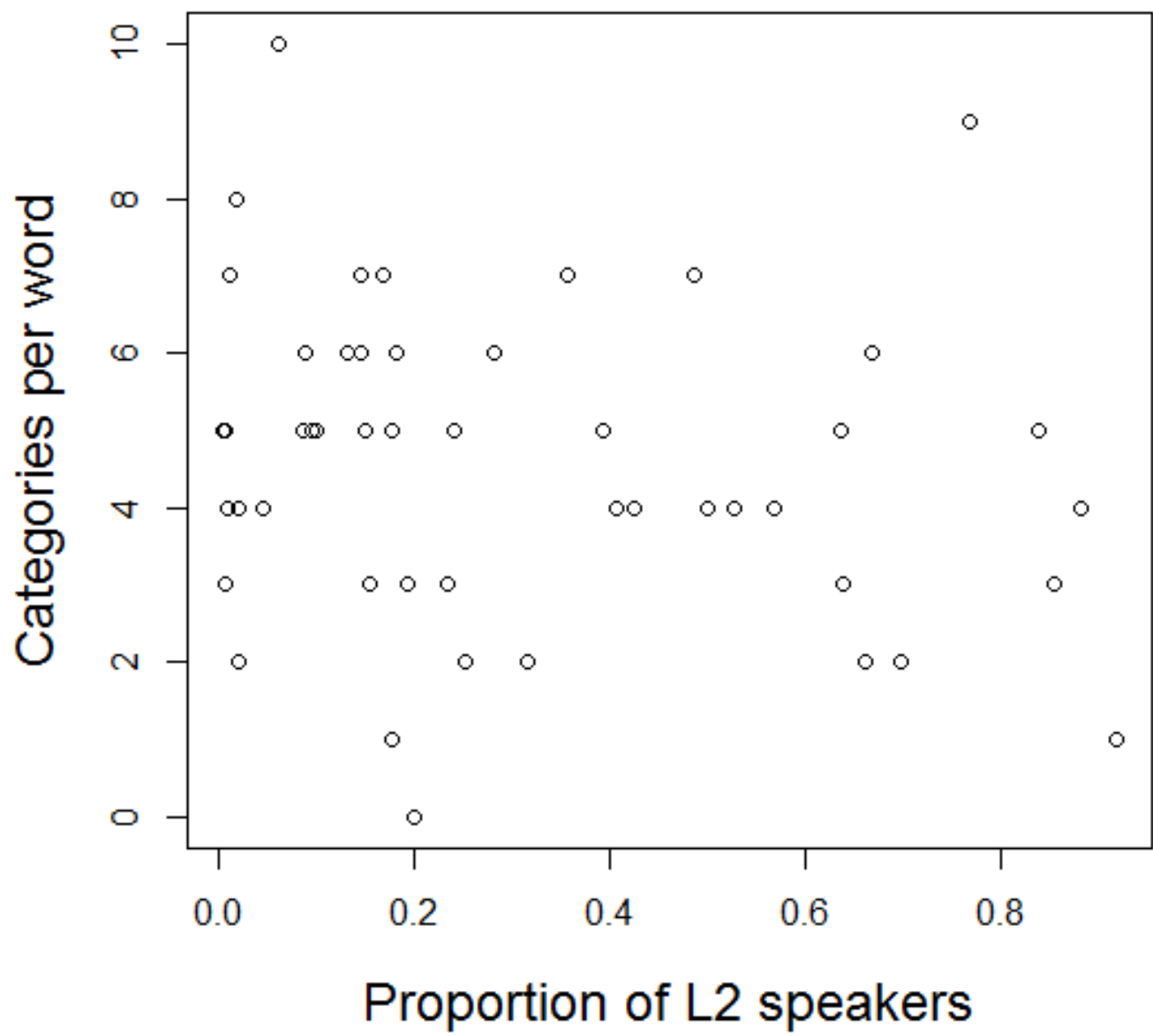


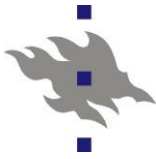
→ Test whether we can do without random slopes.

- If yes, we can assume that a significant main effect is independent of families and areas.
- And leave random intercepts only.



- Families were modeled as stocks (*Autotyp*) and areas as continents (10 areas in the *Autotyp*).
- Each factor's effect was evaluated using likelihood ratio test where a model with the variable of interest was compared to a model without the variable of interest.
- L2 % was modeled as the predictor and inflectional synthesis as the response.

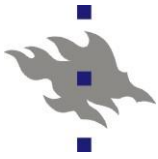




- Area as random slope: $\chi^2 = 0.14$; $df = 2$; $p = 0.93$
- Area as random intercept: $\chi^2 = 4.49$; $df = 1$; $p = 0.034$ (*)
- Stock as random slope: $\chi^2 = 0.05$; $df = 2$; $p = 0.97$
- Stock as random intercept: $\chi^2 = 1.47$; $df = 1$; $p = 0.23$

- L2 %: $\chi^2 = 1.69$; $df = 1$; $p = 0.19$

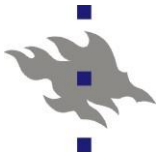
→ The model suggest that the proportion of second language speakers (L2 %) has no effect on the distribution of inflectional synthesis.



On social data

- Data on the number of L1 speakers.
 - Main source Ethnologue (Lewis et al. 2016). Census data problematic.
 - Data even for L1 may be difficult to come by.

- L2 data.
 - Very hard to get data for languages outside Eurasia.
 - Bentz & Winter: ~80% of data from Eurasia & Africa.
 - Figures for small languages may be misleading.
 - When the age of native speakers 40+, the youth learn it as a second language. → Elders unlikely to simplify the group's traditional language (?).

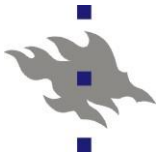


- How is “second language” defined in the sources?
 - How about semi-speakers?
 - How are bilinguals counted?

- E.g. Windfuhr (2008: 418) on Persian:
 - L2 = "at least use or understand Persian as a second language“

- Current sociolinguistics \neq sociolinguistics at the time when the linguistic patterns were formed.
 - Number-driven approach is clearly not optimal.

- Solution: detailed sociolinguistic profiles of languages.

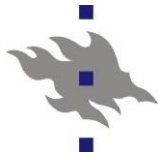


Conclusion

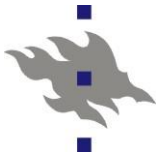
- These preliminary results suggest that
 - the proportion of second language speakers in the speech community has no effect on the degree of inflectional synthesis of the verb.

- I also wonder whether L2 % is methodologically a feasible way to model sociolinguistics of language contact.

- In future research a more promising way would be to survey sociolinguistic profiles of languages in detail.



Thank you!

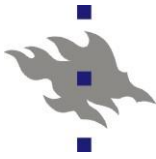


References

- Bentz, C. & B. Winter 2013. Languages with more second language learners tend to lose nominal case. *Language Dynamics and Change* 3, 1-27.
- Bybee, J. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press.
- Dammel, A. & S. Kürschner 2008. Complexity in nominal plural allomorphy: A contrastive survey of ten Germanic languages. In M. Miestamo et al. (eds), 243-262. *Language Complexity: Typology, Contact, Change*. Amsterdam: John Benjamins.
- DeLancey, S. 2014. Sociolinguistic typology in North East India: A tale of two branches. *Journal of South Asian Languages and Linguistics* 1(1): 59-82).
- Givón, T. 2009. *The Genesis of Syntactic Complexity: Diachrony, Ontogeny, Neuro-Cognition, Evolution*. Amsterdam: Benjamins.
- Jaeger T. F. et al. 2011. Mixed effect models for genetic and areal dependencies in linguistic typology. *Linguistic Typology* 15: 281-320.
- Kaye, J. 1989. *Phonology: A cognitive view*. Hillsdale, NJ: Lawrence Erlbaum.
- Kusters, W. 2003. *Linguistic complexity: The influence of social change on verbal inflection*. Ph.D. diss., University of Leiden.



- Labov, W. (1996). *Principles of Linguistic Change, Vol. I: Internal Factors*. Oxford: Blackwell.
- Ladd, D. R. et al. 2015. Correlational studies in typological and historical linguistics. *Annual Review of Linguistics* 1: 221-241.
- Lewis, M. P., G. F. Simons, and C. D. Fennig (eds.) 2016. *Ethnologue: Languages of the World (19th edn)*. Dallas, TX: SIL International. <http://www.ethnologue.com>.
- Lupyan, G. & R. Dale 2010. Language structure is partly determined by social structure. *PLoS One* 5: e8559.
- Nettle, D. 2012. Social scale and structural complexity in human languages. *Philosophical Transactions of the Royal Society B* 367: 1829-1836.
- Newmeyer, F. 2002. Uniformitarian assumptions and language evolution research. In A. Wray (ed.), *The transition to language*, 359-375. Oxford: OUP.
- Nichols, J., A. Witzlack-Makarevich & Balthasar Bickel 2013. The autotyp genealogy and geography database. <http://www.autotyp.uzh.ch>



- Nichols, J., A. Witzlack-Makarevich & Balthasar Bickel (in press).
Inflectional synthesis on the verb. Autotyp Database.
- Parodi, T., B. D. Schwartz, and H. Clahsen 2004. On the L2 acquisition of
the morphosyntax of German nominals. *Linguistics* 42: 669-705.
- Sinnemäki, K. 2010. Word order in zero-marking languages. *Studies in
Language* 34: 869-912.
- Thomason, S. G. 2008. Does Language Contact Simplify Grammars?
Deutsche Gesellschaft für Sprachwissenschaft, Bamberg, February
2008. <http://www-personal.umich.edu/~thomason/temp/simple2.pdf>.
- Trudgill, P. 2011. Sociolinguistic typology: Social determinants of linguistic
complexity. Oxford: OUP.
- Trudgill, P. (to appear). The Anthropological Setting of Polysynthesis. In N.
Evans et al., M. Fortescue & M. Mithun (eds), *The Oxford Handbook of
Polysynthesis*. Oxford: OUP.
- Windfuhr, G. & J. R. Perry 2009. Persian and Tajik. In G. Windfuhr (ed.),
The Iranian Languages, 416-544. London: Routledge.