# Language complexity: Why is it relevant for typologists?

Department of General Linguistics, Zurich, Nov 26, 2013

Kaius Sinnemäki,
Helsinki Collegium for Advanced Studies
University of Helsinki (kaius (dot) sinnemaki (at) helsinki (dot) fi)

## Background

- The notion of complexity is notoriously difficult to define and to harness for measurement (Comrie 1992).

- Despite these problems, typological research on complexity has proliferated during the last 10+ years:
  - Monographs (Nichols 1992, Perkins 1992, Dahl 2004, Hawkins 2004, McWhorter 2007, Givón 2009, Trudgill 2011).
  - Edited volumes (Miestamo et al. 2008, Sampson et al. 2009, Pellegrino et al. 2009, Newmeyer et al. forthcoming).
  - Special journal issues/commentaries (*Linguistic Typology* 5/2-3, 8/3, 15/2, *Science* 335/6069).
  - Dissertations (Kusters 2003, Sinnemäki 2011).

Notwithstanding the titles of this paper [How complex are isolating languages] and this volume [Language complexity: Typology, contact, change], I remain agnostic as to whether the notion of complexity has an important role to play in the study of language, and, in particular, the field of linguistic typology. (In fact, I suspect that it may prove to be more relevant to domains such as phylogeny, diachrony, ontogeny and sociolinguistics than to the ”straight” synchronic study of language.)

David Gil (2008: 129-130)

# But what is typology?

- Evaluation against which conception, or aim, of typology?
  - What is typology? (Bickel 2007; Sinnemäki 2011).
- Classically, linguistic typology seen as a flipside of universal grammar.
  - Main contribution within cognitive sciences.
- In modern research, typology as a discipline of its own.
  - Its own agenda, theories, methods, problems, association, conferences, journals & other publication forums, etc.
  - Main contribution to other human sciences (cognitive studies, cultural studies, archeology, genetics).

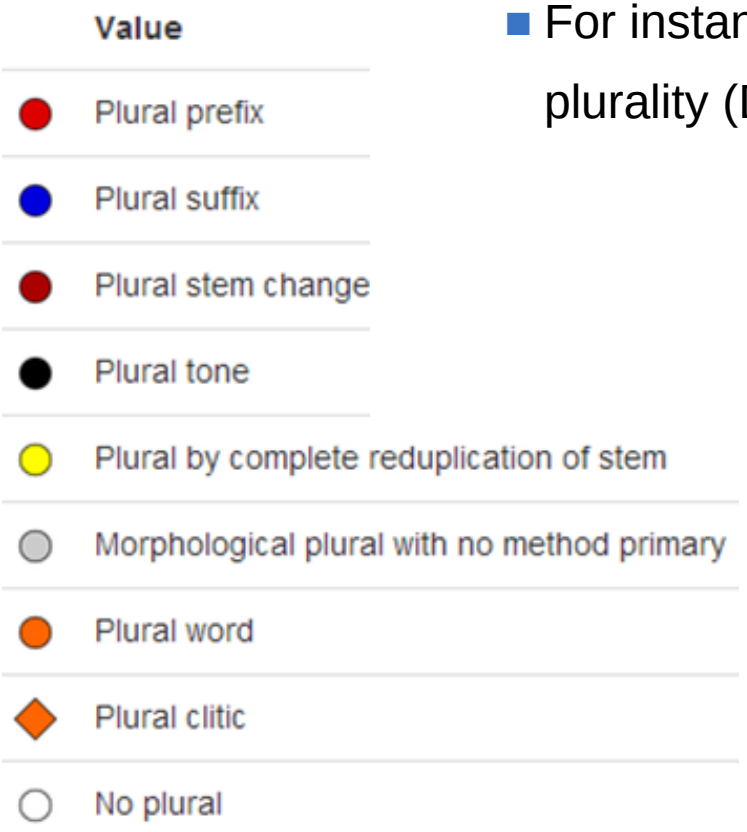
- What are the goals of modern typology? (Bickel 2007)?

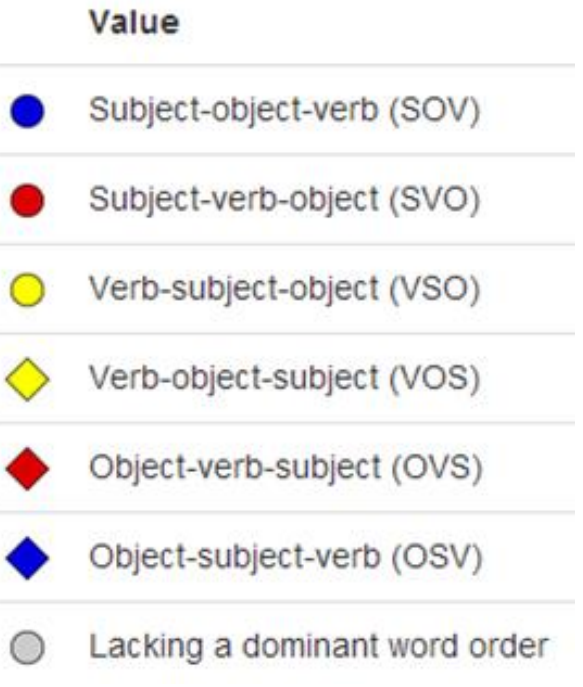
    1. Uncover cross-linguistic diversity and unity based on a wide range of languages.

    2. Investigate whether linguistic patterns interact
        a. among themselves,
        b. with i) cognitive, ii) sociocultural and iii) genetic patterns.

- My aim: to argue that the notion of complexity may help us realize goal #2.

- Why not goal #1?
  - Not all linguistic patterns can be meaningfully described with the notion of complexity
  - For instance word order parameters, coding of nominal plurality (Dryer 2011a,b), etc.



## Plan

- What is complexity?
- Interrelations among linguistic patterns *in terms of complexity* (Sinnemäki 2011, in preparation).
- Typological distributions may reflect varying degrees of processing complexity (Hawkins 2004).
- Typological distribution of complexity may correlate with socio-cultural patterns (Trudgill 2011).
- How about the connection between typological and genetic patterns (Dediu & Ladd 2007)?

# What is complexity?

- Exact definition of complexity is enormously difficult.
  - Mikulecky (2001): that's what its about.
  - Analytic vs. holistic approaches.

- We need several different concepts when approaching complexity (Rescher 1998, Sinnemäki 2011).

- At a general level, complexity can be characterized as the number and variety of elements and the elaborateness of their interrelational structure (Simon 1996, Rescher 1998, Hübler 2007, Sinnemäki 2011, among many others).

- The general characterization can be broken down into more fine-grained characterizations (Rescher 1998, Moravcsik and Wirth 1986)
  - Number of parts: syntagmatic complexity.
    - Word length in terms of phonemes, syllables, etc. (Nettle 1995; Fen-Oczlon & Fenk 1999; Wichmann et al. 2011).
    - Constituent length (Bresnan et al. 2007; Diessel 2008).
  - Variety of parts: paradigmatic complexity.
    - Phoneme inventory size (Shosted 2006).
    - Number of distinctions in a grammatical category, e.g. aspect (McWhorter 2001).

- Organizational (or network?) complexity:
  - Ways of arranging components in different modes of interrelationship.
  - E.g., phonotactic restrictions, variety of distinctive word orders (Sinnemäki, forthcoming).

- Hierarchic complexity.
  - Recursion (Chomsky 1965; Givón 2009), lexical-semantic hierarchies, …

- While the parts of a system are necessary building blocks of complexity, it is interactions that really characterize the degree of complexity of a system or its subpart.

| General characterization of complexity | Number and variety of parts and their interrelations | | | |
|---|---|---|---|---|
| Detailed characterization of complexity | Syntagmatic complexity | Paradigmatic complexity | Organizational complexity | Hierarchical complexity |
| Measures of complexity | Entropy, size, description length, effective complexity, information, connectivity, irreducibility, low probability, syntactic depth, … | | | |
| Linguistic analysis | Typology, classification | | | |
| Data | Texts, grammatical descriptions, … | | | |

- Complexity can be related also to usage or operation of a system (operational complexity; Rescher 1998).
  - E.g. cost-related differences concerning production and comprehension (Gibson 1998), efficiency (Hawkins 2004).

- In the typological literature usage-complexity is commonly separated from the descriptive complexity of linguistic patterns (Dahl 2004, Miestamo 2008).
  - But see Kusters (2003).

# Interactions among linguistic patterns

- Interactions in terms of complexity have been assumed to be balancing, so that if one pattern is complex, another is likely to be simple – and vice versa.

  …impressionistically it would seem that the total grammatical complexity of any language…is about the same as that of any other. This is not surprising, since all languages have about equally complex jobs to do…

  Charles Hockett (1958: 180-181)

- How much is it actually about functional load?

  "The notion of functional load is that a phonemic system L has a (quantifiable) job to do" (Hockett 1966: 8).

- Such complexity trade-offs do occur (Sinnemäki 2011, forthcoming, pace e.g. Shosted 2006).

- A few examples here, others discussed in Sinnemäki (in preparation).

# Locus of marking in possessive NPs (Nichols & Bickel 2005)

Head marking

*Bonnie ø-avhay*

Bonnie 3-dress

‘Bonnie’s dress’

(Maricopa, Gordon 1986: 31)

Double marking

*heidä-n talo-nsa*

3PL-POSS house-3.POSS

‘their house’

(Finnish, Karlsson 1999: 98)

Dependent marking

*hakew-kate tahu*

Raquel-POSS knife

‘Raquel’s knife’

(Trumai, Guirardello 1999: 76)

Zero marking

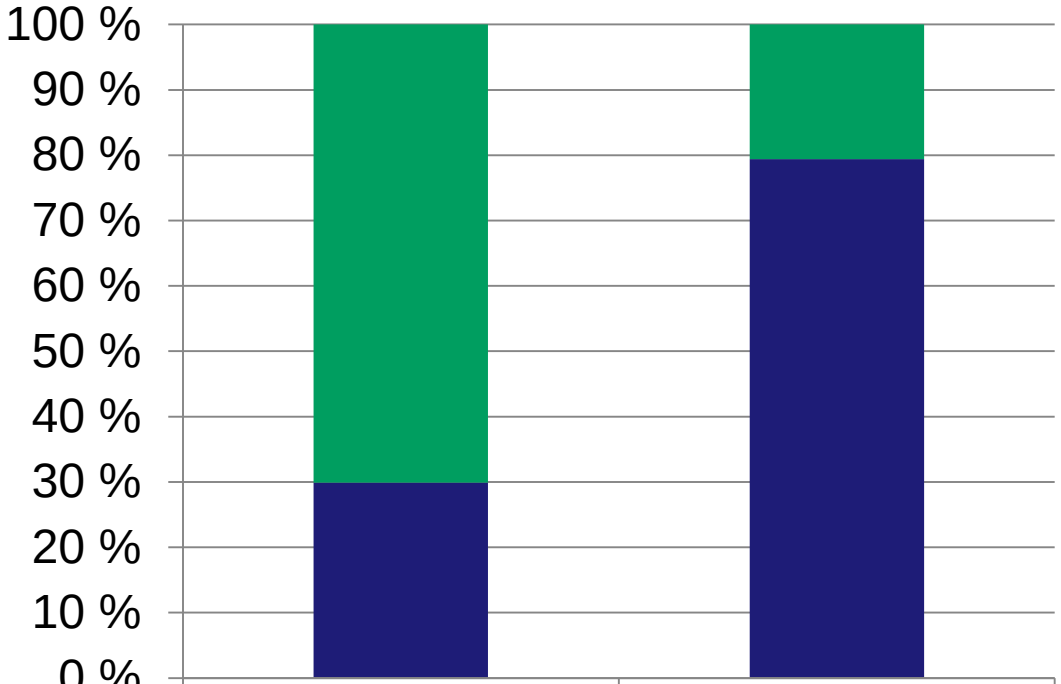
*rumah Tomo*

house Tomo

‘Tomo’s house’

(Indonesian, Sneddon 1996: 144)

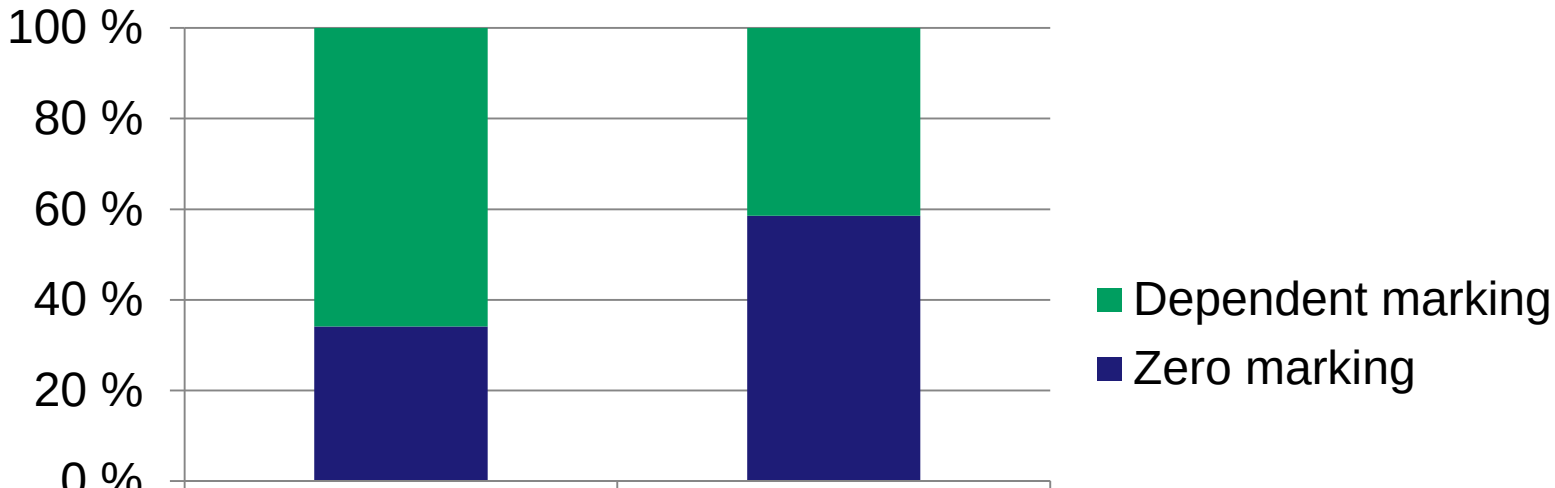
# Locus of marking in possessive NPs




| | Zero marking | Head marking |
|---|---|---|
| Dependent marking | 75 | 21 |
| Zero marking | 32 | 81 |

- Negative correlation : tau-a = -.50; p < .0001 (counting genera). Regression modeling lead to similar results.

# Locus of marking in the clause, 1st and 2nd person pronouns (Comrie 2005, Siewierska 2005)

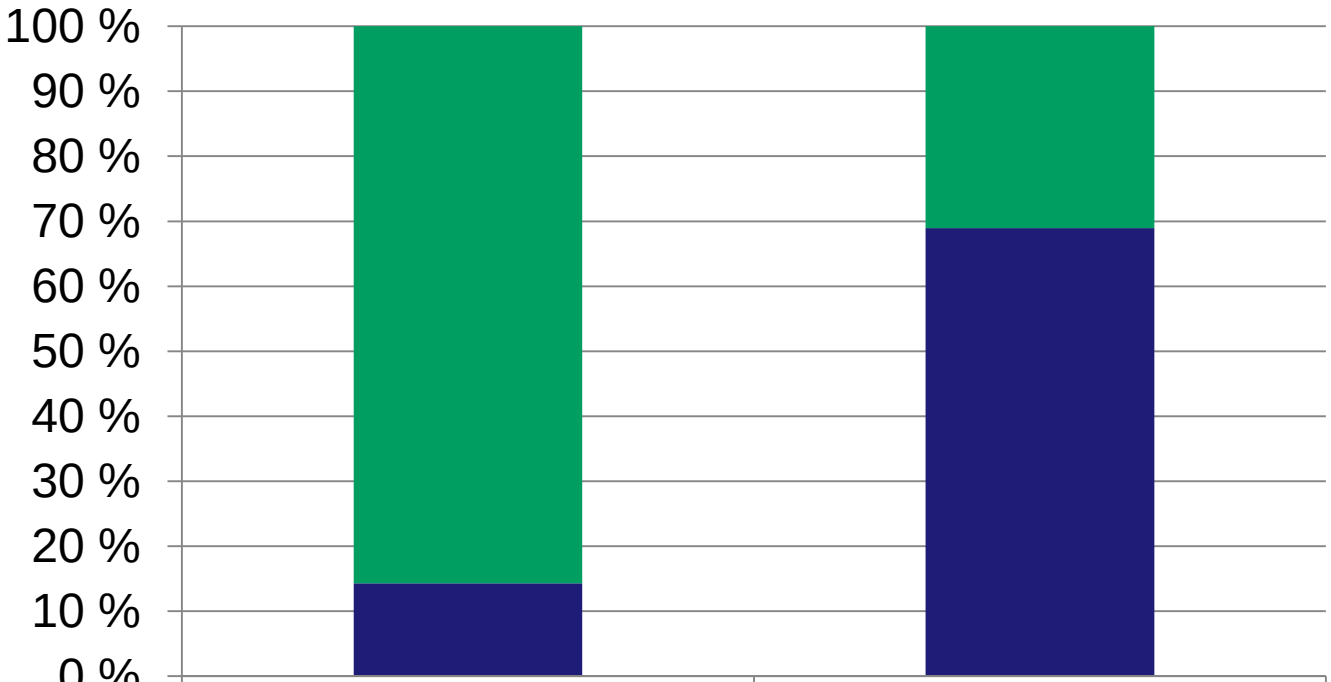



| | Zero marking | Head marking |
|---|---|---|
| Dependent marking | 56 | 29 |
| Zero marking | 29 | 41 |

- Combining data from Comrie (2005), alignment of case marking in pronouns and Siewierska (2005) on alignment of person marking in the verb (counting genera).

- Negative correlation: tau-a = -.24, p = .0024.

# Case marking and rigid order in core argument marking

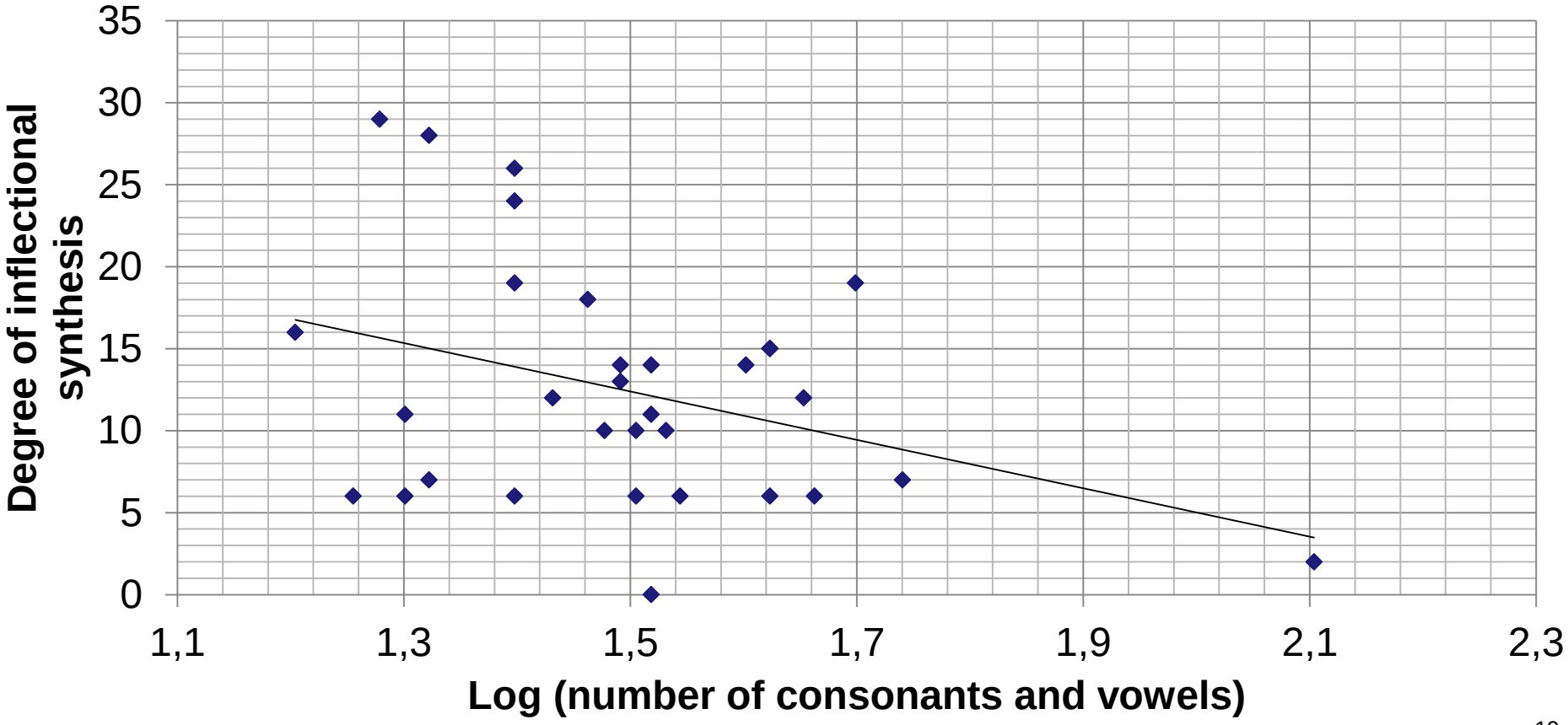



| | free order | rigid order |
|---|---|---|
| case marking | 18 | 9 |
| zero marking | 3 | 20 |

- Negative correlation: tau a = -.54, p = .0002 (Sinnemäki, forthcoming).

# Phoneme inventory and inflectional synthesis (Shosted 2006)

- Negative correlation (Pearson's r = -.36, p < .05). Data on inflectional synthesis from Bickel and Nichols 2005).



# Equi-complexity

- A basic assumption in language complexity research:
  - If languages were equally complex, then complexity trade-offs would need to be an all-encompassing principle in language.
- Claim 1: a handful of correlation-pairs cannot validate (or falsify) the equi-complexity hypothesis.
- Claim 2: these hypotheses need to be kept separate.
  - Hypothesis 1 (equi-complexity hypothesis): Languages are roughly equal in their overall complexity.
  - Hypothesis 2 (trade-off hypothesis): Complexity in one area of grammar correlates with simplicity in another.

# Simulating typological data

- Let's assume we have data on 300 "languages", 40 variables, and we know their complexity values.
  - The range of the values roughly mimic that in *The WALS*: in most languages from 0 and 1, maximally from 0 to 7.

- Let the complexity values vary randomly but keeping the "sum of complexity" constant → equal complexity.

- Let's form correlation pairs between all the variables, use Kendall's tau-a, and permute the data 1000 times.

- Altogether 780 correlation pairs. On average
  - 6.1 % negative ($p < .05$). correlations (random: 2.5%).
  - 1.3 % positive ($p < .05$) correlations (random: 2.5%).
  - About 92% of the correlations were non-significant.

→ The likelihood of finding a complexity trade-off is quite small even if the data simulated equal complexity!

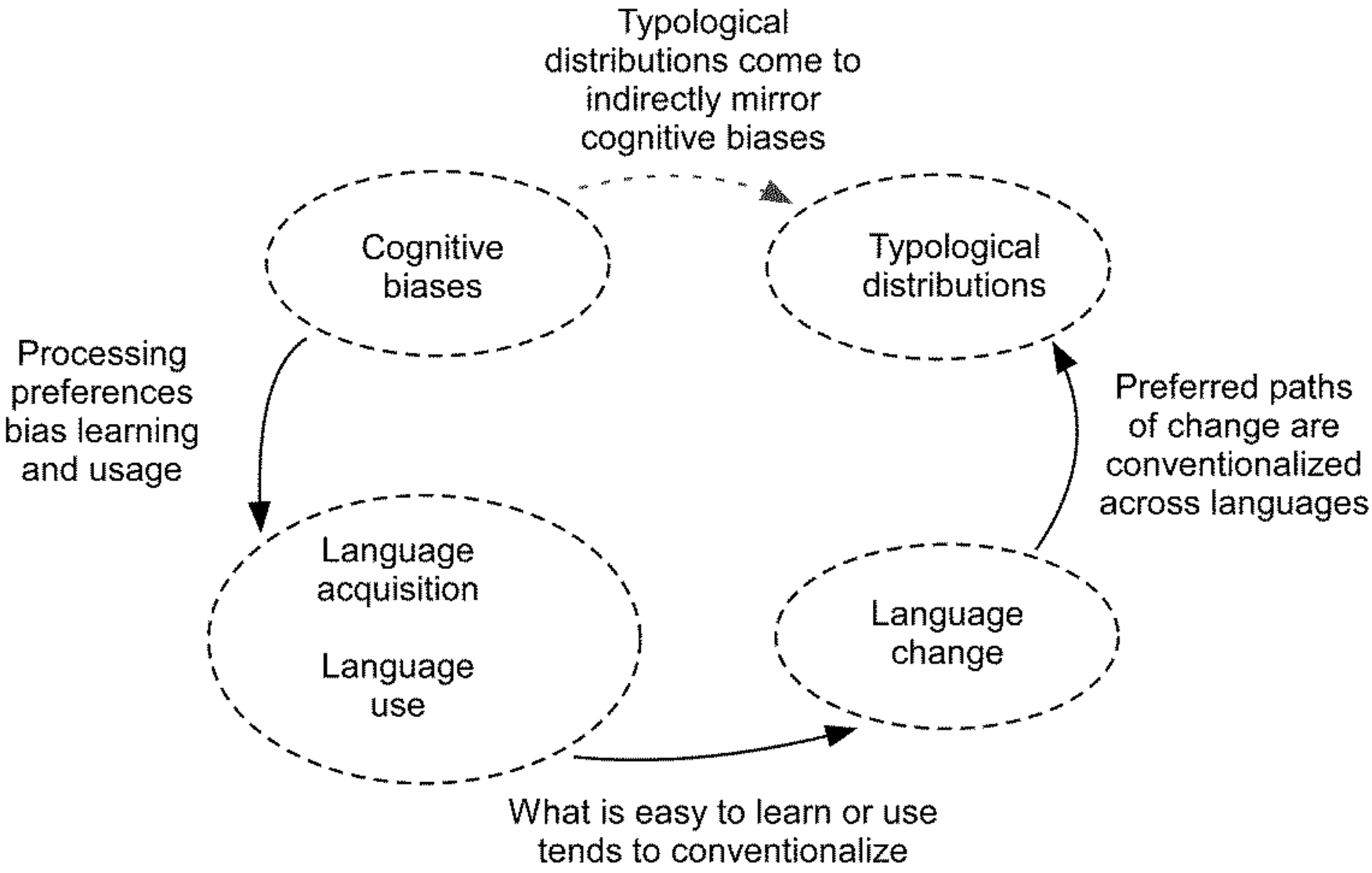
→ But: In all simulations, there were at least twice as many significant negative than positive correlations.
  - We would need to study interactions between variables, but this is difficult given the typological data.

- How about the WALS data (Haspelmath et al. 2005)?
  - 40 features with complexity interpretation.
  - Altogether 780 correlation pairs. On average
  - 10.6 % positive ($p < .05$). correlations (random: 2.5%).
  - 7.2 % negative ($p < .05$) correlations (random: 2.5%).
  - More positive than negative correlations!

# Interactions with cognitive patterns

- Very little studied. Only 1-2 experiments on Greenbergian word order correlations (Sinnemäki, submitted).

- How does complexity help here? Two possibilities.
  - Complexity trade-offs point to balancing effect in cognition.
    - Distinctness and economy in processing (Bornkessel-Schlesewsky and Schlesewsky 2009).
  - Focusing on processing complexity, that is, the relative ease vs. difficulty of processing or acquiring a linguistic pattern.
    - What is easy to learn or use, tends to get grammaticalized and ends up in typological preferences.


Typological distributions come to indirectly mirror cognitive biases
Cognitive biases
Typological distributions
Processing preferences bias learning and usage
Preferred paths of change are conventionalized across languages
Language acquisition
Language use
Language change
What is easy to learn or use tends to conventionalize

# Examples

- A lot of experimental evidence for the subject preference in declarative clauses and relative clauses.
  - E.g., SO order easier to process than OS order. Easier to process subject relative clauses than object relative clauses.

    (a) The man that $___{SUBJ}$ drove the car.

    (b) The car that the man drove $___{OBJ}$.

  - But, e.g. in Basque object relative clauses appear easier to process (Carreiras et al. 2010).

- At least some support for the following:
  - Case marking and rigid word order.
  - SOV and case marking.
  - SVO and zero marking.
  - Differential object marking affected by animacy.
  - Suffixes preferred over prefixes.
  - Greenbergian word order correlations (OV/VO, prep/postp, N-gen/gen-N).

- Point to consider:
  - 50 years since Greenberg (1963), but a single experiment in a conference proceedings to test OV/VO, position of adpositions and N-Gen/Gen-N (Christiansen 2000).

- Subject preference fairly well-researched with different experimental methods in European languages.

- Other typological preferences experimented mostly using artificial language learning (Culbertson 2012).
  - The subjects are taught miniature languages and their performance is then tested, how they change the language or how they extend the patterns to new data.
  - Problems:
    - Simulates second language learning by adults.
    - Has not been replicated in iterated learning experiments (Rafferty et al. 2013).

# Interactions with sociocultural patterns

- Basic question: do the properties of the speech community affect language structure.

- Increasing evidence that it does.
  - Kusters (2003): verbal inflection simplifies if the language is learned as second language by many adults.
  - Peter Trudgill has argued for at least ten years that the degree of language contact, size and network structure may affect the way language structure simplifies or becomes more complex (2011).
  - Lupyan & Dale (2010) use the WALS data to argue for this.

- What is the rationale behind this?

- Cognitive research has overlooked sociocultural differences in cognition (Levinson & Gray 2012).
  - Because languages are learned and used in different sociocultural environments, those contexts may bias language use and learning so much so that language structure is adapted to those biases.
  - This may end up reflected in typological distributions as well (Sinnemäki 2009, Lupyan & Dale 2010).

- Problem: chance correlations (Roberts & Winters 2012).
  - N of basic color categories correlates with the number of colors in the national flag ($r = 0.12$, $p = .03$).
  - Countries with acacia tree (Acacia nilotica) are more likely to have tonal languages ($\chi^2 = 47.1$, $p < .0001$).
  - Basic word order of a language predicts how many children a couple will have ($F(3,35030) = 121.6$, $p < 0.0000001$), accounted for 36% of variation in the n of children.

- Need for a better articulated theory and the use of methods that allow finding dynamic universals.

# Conclusion

- The notion of complexity may not increase our ability to devise more fine-grained typological classifications.
- There is some evidence for interrelations among linguistic patterns in terms of complexity.
  - Do all of them point to balancing effects?
- Complexity is useful for elucidating the interrelations between linguistic patterns on the one hand , and cognitive and sociocultural patterns, on the other.

Thank you!

# References


Bickel, B. & J. Nichols 2005. Inflectional synthesis of the verb. In M. Haspelmath, M. S. Dryer, D. Gil & B. Comrie (eds.), The world atlas of language structures, 94-97. Oxford: OUP.

Bickel, B. 2007. Typology in the 21st century: Major current developments. Linguistic Typology 11: 239-251.

Bornkessel-Schlesewsky, I. & M. Schlesewsky 2009. The role of prominence information in the real-time comprehension of transitive constructions: a cross-linguistic approach. Language and Linguistics Compass 3: 19-58.

Bresnan, J, A. Cueni, T. Nikitina & R. H. Baayen 2007. Predicting the Dative Alternation. In G. Boume, I. Kraemer & J. Zwarts, Cognitive foundations of interpretation, 69-94. Amsterdam: Royal Netherlands Academy of Science.

Carreiras, M., J. A. Duñabeitia, M. Vergara, I. de la Cruz-Pavía, I. Laka 2010. Subject relative clauses are not universally easier to process: evidence from Basque. Cognition 115: 79-92.

Chomsky, N. 1965. Aspects of the theory of syntax. Cambridge, MA: MIT Press.

Christiansen, M. H. 2000. Using artificial language learning to study language evolution: Exploring the emergence of word order universals. In J. L. Dessalles & L. Ghadakpour (eds), The evolution of language: 3rd International Conference, 45-48. Paris: Ecole Nationale Superieure des Telecommunications.

Comrie, B. 1992. Before complexity. In J. A. Hawkins & M. Gell-Mann (eds.), The evolution of human languages, 193-211. Redwood City: Addison-Wesley.

Comrie, B. 2005. Alignment of case marking. In M. Haspelmath, M. S. Dryer, D. Gil & B. Comrie (eds.), The world atlas of language structures, 398–405. Oxford: OUP.

Culbertson, J. 2012. Typological universals as reflections of biased learning: evidence from artificial language learning. Language and Linguistics Compass 6: 310-329.

Dahl, Ö. 2004. The growth and maintenance of linguistic complexity. Amsterdam: Benjamins.

Dediu D., D. R. Ladd 2007. Linguistic tone is related to the population frequency of the adaptive haplogroups of two brain size genes, ASPM and Microcephalin. PNAS 104: 10944-10949.

Diessel, H. 2008. Iconicity of sequence: A corpus-based analysis of the positioning of temporal adverbial clauses in English. Cognitive Linguistics 19: 465-490.
Dryer, M. S. 2011a. Coding of nominal plurality. In: M. Dryer & M. Haspelmath (eds.), The world atlas of language structures online. Leipzig: Max Planck Institute for Evolutionary Anthropology. (http://wals.info/chapter/33.)
Dryer, M. S. 2011b. Order of subject, object and verb. In: M. Dryer & M. Haspelmath (eds.), The world atlas of language structures online. Leipzig: Max Planck Institute for Evolutionary Anthropology. (http://wals.info/chapter/81.)
Fenk-Oczlon, G. & A. Fenk 1999. Cognition, quantitative linguistics, and systemic typology. Linguistic Typology 3: 151-177.
Gibson, E. 1998. Linguistic complexity: Locality of syntactic dependencies. Cognition 68: 1-76.
Gil, D. 2008. How complex are isolating languages. In M. Miestamo, K. Sinnemäki & F. Karlsson (eds.), Language complexity: Typology, contact, change, 109-131. Amsterdam: Benjamins.
Givón, T. 2009. The genesis of syntactic complexity: Diachrony, ontogeny, neuro-cognition, evolution. Amsterdam: Benjamins.

Gordon, L. 1986. Maricopa morphology and syntax. Berkeley, CA: University of California Press.
Greenberg, J. H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In J. H. Greenberg (ed.), Universals of grammar, 73-113. Cambridge, MA: MIT Press.
Guirardello, R. 1999. A reference grammar of Trumai. Houston, TX: Rice University dissertation.
Haspelmath, M., M. S. Dryer, D. Gil & B. Comrie (eds.) 2005. The world atlas of language structures. Oxford: OUP.
Hawkins, J. A. 2004. Efficiency and complexity in grammars. Oxford: OUP.
Hockett, C. F. 1958. A course in modern linguistics. New York, NY: Macmillan.
Hockett, C. F. 1963. The quantification of functional load: A linguistic problem. Santa Monica: Rand Corp.
Hübler, A. W. 2007. Understanding complex systems: Defining an abstract concept. Complexity 12: 9-11.
Karlsson, F. 1999. Finnish: An essential grammar. London: Routledge.
Kusters, W. 2003. Linguistic complexity: The influence of social change on verbal inflection. University of Leiden dissertation.

Levinson, S. C. & R. D. Gray 2012. Tools from evolutionary biology shed new light on the diversification of languages. Trends in Cognitive Sciences 16: 167-173.
Lupyan, G. & R. Dale 2010. Language structure is partly determined by social structure. PLoS One 5: e8559.
McWhorter, J. 2001. The world's simplest grammars are creole grammars. Linguistic Typology 5: 125-156.
McWhorter, J. 2007. Language interrupted: Signs of non-native acquisition in standard language grammars. Oxford: OUP.
Miestamo, M. 2008. Grammatical complexity in a cross-linguistic perspective. In M. Miestamo, K. Sinnemäki & F. Karlsson (eds.), Language complexity: Typology, contact, change, 23–41. Amsterdam: Benjamins.
Miestamo, M., K. Sinnemäki & F. Karlsson (eds.) 2008. Language complexity: Typology, contact, change. Amsterdam: Benjamins.
Mikulecky, D. C. 2001. The emergence of complexity: Science coming of age or science growing old. Computers & Chemistry 25: 341-348.
Moravcsik, E. A. & J. R. Wirth 1986. Markedness: An overview. In F. R. Eckman, E. A. Moravcsik & J. R. Wirth (eds.), Markedness, 1-11. New York: Plenum.

Nettle, D. 1995. Segmental inventory size, word length, and communicative efficiency. Linguistics 33: 359-367.
Newmeyer, F. J. & L. B. Preston (eds.) (forthcoming). Measuring linguistic complexity. Oxford: OUP.
Nichols, J. & B. Bickel 2005. Locus of marking in the clause; Locus of marking in possessive noun phrases; Locus of marking: Whole-language typology. In M. Haspelmath, M. S. Dryer, D. Gil & B. Comrie (eds.), The world atlas of language structures, 98-109. Oxford: OUP.
Nichols, J. 1992. Linguistic diversity in space and time. Chicago: The University of Chicago Press.
Pellegrino, F., E. Marsico, I. Chitoran & C. Coupé (eds.) 2009. Approaches to phonological complexity. Berlin: Mouton de Gruyter.
Perkins, R. D. 1992. Deixis, grammar and culture. Amsterdam: Benjamins.
Rafferty, A. N., T. L. Griffiths & M. Ettlinger 2013. Greater learnability is not sufficient to produce cultural universals. Cognition 129: 70-87.
Rescher, N. 1998. *Complexity: A philosophical overview*. New Brunswick: Transaction.
Roberts, S. & J. Winters 2012. Social structure and language structure: The new nomothetic approach. Psychology of Language and Communication 16: 89-112.

Sampson, G., D. Gil & P. Trudgill (eds.) 2009. Language complexity as an evolving variable. Oxford: OUP.
Shosted, R. 2006. Correlating complexity: A typological approach. Linguistic Typology 10: 1-40.
Siewierska, A. 2005. Verbal person marking. In M. Haspelmath, M. S. Dryer, D. Gil & B. Comrie (eds.), The world atlas of language structures, 414–417. Oxford: OUP.
Simon, H. A. 1996. The sciences of the artificial (3rd edn). Cambridge, MA: MIT Press.
Sinnemäki, K. 2009. Complexity in core argument marking and population size. In G. Sampson, D. Gil & P. Trudgill (eds), Language complexity as an evolving variable, 125-140. Oxford: OUP.
Sinnemäki, K. 2011. Language universals and linguistic complexity: Three case studies in core argument marking. University of Helsinki dissertation. http://urn.fi/URN:ISBN:978-952-10-7259-8.
Sinnemäki, K. (forthcoming). Complexity trade-offs: A case study. In F. J. Newmeyer & L. B. Preston (eds.), Measuring linguistic complexity. Oxford: OUP.
Sinnemäki, K. (in preparation). Complexity trade-offs: A typological survey.
Sinnemäki, K. (submitted). Processing complexity, typology, variation.

Sneddon, J. N. 1996. Indonesian: A comprehensive grammar. London: Routledge.
Trudgill, P. 2011. Sociolinguistic typology: Social determinants of linguistic complexity. Oxford: OUP.
Wichmann, S., T. Rama & E. W. Holman 2011. Phonological diversity, word length, and population sizes across languages: The ASJP evidence. Linguistic Typology 15: 177-197.