Linguistic complexity: Possibilities and limitations for a typologist

Department of Linguistics
Johannes Gutenberg University of Mainz
June 8, 2015

Kaius Sinnemäki,
Helsinki Collegium for Advanced Studies
University of Helsinki (kaius.sinnemaki@helsinki.fi)
1. Background

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

- Despite this problem, cross-linguistic research on complexity has proliferated during the last 10+ years:
  - Dissertations (Kusters 2003, Sinnemäki 2011).
Two common objections & answers

1. There is no objective way of defining complexity (e.g. Ansaldo & Nordhoff 2009: 345–346).

   ▪ Instead of searching for an objective metric of complexity, we should understand its diverse nature (objection #1).

2. “…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.” (Gil 2008: 129-130)

   ▪ The notion of complexity may help realize some central goals of language typology (objection #2).
2. What is complexity?

- Exact definition of complexity is enormously difficult.
  - Mikulecky (2001): Complexity is “the property of a real world system that is manifest in the inability of any one formalism being adequate to capture all its properties.”

  ➔ Focus should not be in searching for the right way of defining and measuring complexity.

  - Different concepts needed, some more useful than others for different purposes (Rescher 1998, Sinnemäki 2011).
  - Roberts & Bresnan (2008):
    - Inflectional vs. contextual inflection in pidgins.
At a general level, complexity characterized as the number and variety of elements and their interactions (Simon 1996, Rescher 1998, Hübler 2007, among others).

Strong consensus about this!

This characterization broken down into more fine-grained classifications (Rescher 1998, Moravcsik and Wirth 1986).

See Sinnemäki (2011) for more.

Variety of parts: paradigmatic complexity.

Number of cases in the system:
- Finnish about 15, German four, Mandarin none.
Organizational (or network) complexity:

- Ways of arranging parts in different kinds of relationships.
- Does case interact with other grammatical systems?
  - DOM: case tends to depend on definiteness, tense, animacy or number (Sinnemäki 2014a, 2014b).
  - Example from Kashmiri (Bhatt 1999: 38, 41).

\[ Maastar \quad laa-yi \quad lar\-k-as. \]
\[ \text{teacher.NOM beat-FUT.3SG boy-DAT} \]

‘The teacher will beat the boy.’

\[ Lark-an \quad khy-av \quad bat\-i. \]
\[ \text{boy-ERG ate-M.SG food.M.SG.NOM} \]

‘The boy ate the food.’
Usage complexity (or difficulty):

- Acquisition, learning and processing.
  - Involves measuring e.g. reaction times, brain activity, or behavioral preferences.
  - Difficulty related to unexpectedness and/or memory capacity (Levy et al. 2013).

- Case marking (in particular) tends to be difficult for adult second language learners (see Bentz & Winter 2013: 3-4 & Trudgill 2011 and references there).
Two common dimensions of language complexity (cf. Kusters 2003; Miestamo 2008).

- Economy (both paradigmatic and syntagmatic):
  - number and variety of parts.

- Transparency (both paradigmatic and syntagmatic):
  - Adherence to / deviation from one-meaning-one-form.

Consequences to the study of language complexity:

- When studying the complexity of a particular linguistic structure, focus should be
  1) on particular type of complexity, not overall complexity,
  2) in local context, not in all kinds of contexts.
3. 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.

Notion of complexity may help us realize goal #2.

Maybe in a limited way goal #1 (e.g. Stump & Finkel 2013 on morphological typology).
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.
4. Typological correlations

- Interactions in terms of complexity have been assumed to be balancing in language: if one pattern is complex, another is simple – and vice versa (e.g. Hockett 1958).

- Such trade-offs occur in language (pace Shosted 2006).
  - In well-motivated places, e.g. between coding strategies in the same functional domain (Sinnemäki 2011).
  - Not overall (Sinnemäki 2014c).
Example 1: Head and dependent marking

- Head and dependent marking:
  - alternative patterns for marking syntactic relations in different constructions (Nichols 1992).

- In head marking (HM), the syntactic relation marked on the head, in dependent marking (DM) on the dependent.

  *Sabidô ya-wi*
  Sabino  GEN-house
  ‘Sabino’s house.’ (Jones & Jones 1991: 4)
Data and methods

- Possessive NPs, 230 lgs (Nichols & Bickel 2013a).

- Noun arguments of a transitive verb (P-argument), 233 lgs (Nichols & Bickel 2013b).

- Marking of 1\textsuperscript{st} and 2\textsuperscript{nd} person indexes:
  - Case marking of pronouns (Comrie 2013) & person marking on the verb (A & P; Siewierska 2013), 170 lgs.

- Grammatical analyses follow directly from the sources.

- Mixed effects modeling, compare nested models with likelihood ratio test (Bentz & Winter 2013).
The effect of DM significant ($\chi^2 = 14.3; p < .0002$). The correlation estimate was negative (-2.3 ± .4).

No significant effect from areas ($\chi^2 = 7.4; p = .06$) or families ($\chi^2 = 0; p = 1$).
No effect for DM ($\chi^2 = 2.5; p = .12$), but the correlation was negative ($-0.63 \pm 0.4$).

No effect from families ($\chi^2 = 0; p = 0.8$), but a significant effect from areas ($\chi^2 = 18.1, p < 0.0005$).
The effect of DM was significant ($\chi^2 = 7.6; p < .006$), and the correlation was negative (-1.6 ± 0.5).

No effect from families ($\chi^2 = 1.0; p = .99$) or areas ($\chi^2 = 7.3, p = .06$).
Example 2: Case marking and rigid order in core argument marking

Strong negative correlation: $\tau_a = -0.54$, $p = 0.0002$ (Sinnemäki 2014b).
5. Sociolinguistic typology

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

Research on these questions has grown rapidly.
- Now reviews available (Nettle 2012 and Ladd et al. 2015).
Preliminaries

- Sociolinguistic factors → language change.

- Language contact
  - Adult L2 may lead to morphological simplification

- Group size (number of speakers)
  - Small communities may have tight social networks and because of that there’s a lot of shared information.
  - In large communities there is more occasions for communication between strangers.
  - Communication between intimates vs. strangers may lead to linguistic preferences (Givón 2009; Trudgill 2011).
Speech community size

- Lupyan & Dale (2010):
  - WALS features correlated with number of L1 speakers.
  - Inverse relationships:
    - large language $\rightarrow$ low morphological complexity
    - small language $\rightarrow$ high morphological complexity

- Problems:
  - What is the causal link between community size and language structure? Proxy for contact effects.
L2 speakers

- Bentz & Winter (2013):
  - studied the proportion of L2 and its relationship with the presence or number of cases in a language.

\[
\text{prop.L2} = \frac{L2}{L1 + L2}
\]

- Inverse relationship:
  - large proportion of L2 \(\rightarrow\) few cases
  - small proportion of L2 \(\rightarrow\) many cases
Figure 2. (a) Presence and absence of case as a function of L2 speaker proportion. For better visibility, presence and absence points are shown with some random jitter along the y-axis. The curve indicates the fit of the logistic model, which represents the estimated probability of observing a language with case. (b) Case rank as a function of L2 speaker proportion. The curve indicates the fit of the negative binomial model, which represents the estimated number of nominal cases.
Problems:

- Case + word order (verb-final) not controlled for.
- What do the current numbers for the L2 population reflect?
  - Kutenai: L1: 12 (4%) L2: 310 (96%)
  - Ethnic Kutenai trying to learn the elders’ language.
    - Figures do not reflect e.g. “foreigner-directed talk”.
  - Simplification when the main function of language is communicative, not symbolic (Kusters 2003).

Answer:

- more detailed data needed on sociolinguistic situations and social histories.
6. Cognitive processing and typological generalizations

How could complexity help here? Two possibilities.

- Complexity trade-offs point to balancing effects in cognition.
  - Distinctness and economy in processing (Bornkessel-Schlesewsky and Schlesewsky 2009).

- Cognitive processing measured as the relative ease vs. difficulty of processing or acquiring a linguistic pattern.
  - What is easy to learn/use, tends to grammaticalize and end up in typological preferences (Sinnemäki 2014d).
Typological distributions come to indirectly mirror cognitive biases.

Processing preferences bias learning and usage.

Types of change are conventionalized across languages.

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 ___\text{SUBJ} drove the car.
  (b) The car that the man drove ___\text{OBJ}.

- But, e.g. in Basque object relative clauses appear easier to process (Carreiras et al. 2010).
Some support for the following (Sinnemäki 2014d):

- 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 (1966), but a single experiment in a conference proceedings to test OV/VO, position of adpositions and N-Gen/Gen-N (Christiansen 2000).
7. Conclusion

- It is in the nature of complexity to be hard to measure.
  - There is no point in searching for an objective overall measure, or in lamenting about its absence.

- The notion of complexity may help us realize some central goals of language typology.
  - Interactions between grammatical systems.
  - Interaction of grammar and social environment.
  - Processing biases underlying typological generalizations.
Thank you!
References


