

INTRODUCTION: THE PROBLEM OF LANGUAGE

COMPLEXITY

FRED KARLSSON, MATTI MIESTAMO, KAIUS SINNEMÄKI

University of Helsinki

1. Language complexity

The topic of language complexity certainly is a timely one. During the past few years it has surfaced in many different contexts: as one of the themes in the pioneering agenda of the Santa Fe Institute founded by Murray Gell-Mann; in Perkins' (1992) study of the complexity of deictic systems in the languages of the world; in Nichols' (1992) profound analysis of global morphological and syntactic diversity; in Hawkins' (1994, 2004) work on the processing-related foundations of grammar; in Deutscher's (2000) book discussing the evolution of sentential complements in Akkadian; in Kusters' PhD (2003) dissertation on the ultimately social bases of complexity differences and language change; in the extensive discussion on the complexity or simplicity of Creoles initiated by McWhorter (1998, 2001); in Dahl's (2004) book on the origin and maintenance of language complexity; in Chipere's (2003) investigations concerning individual differences in

native language proficiency; in Everett's (2005) fieldwork and interpretations of the low complexity of the Amazonian Pirahã language; in Haspelmath's (2006) proposal that the concept of markedness can be dispensed with in linguistic theory and replaced as an explanatory principle by more fundamental, substantive factors such as phonetic difficulty and pragmatic inferences; in Karlsson's (2007a, b) work on constraints on syntactic embedding complexity; in the 2007 Winter Meeting of the Linguistic Society of America, where a workshop was devoted to the complexity topic; in the "Workshop on Language Complexity as an Evolving Variable" arranged by the Max Planck Institute of Evolutionary Anthropology, Leipzig, in April 2007; in the "Recursion in Human Languages" conference at Illinois State University, Normal, Illinois, in April 2007; and so forth.

Several comprehensive complexity-related research projects are currently under way or have recently been completed. One of them is the project "The Grammatical Complexity of Natural Languages", sponsored by the Academy of Finland for the years 2003-2006 and located at the Department of General Linguistics, University of Helsinki. This project, the Department of General Linguistics, and the Linguistic Association of Finland jointly organized the conference "Approaches to Complexity in Language" in Helsinki on August 24-26, 2005. The current volume contains a thematic selection of the papers read at that conference.

The subtitle of this volume is *Typology, Contact, Change*. These catchwords describe the commonalities of the present papers. Language contact, especially when extensive L2 learning is involved, is a main source of complexity reduction (grammar simplification). By definition, such processes involve language change. But complexity reduction is actually at the heart of many types of language change, especially in morphology and syntax. Contact-induced grammatical change is likely to produce outcomes simpler (in some sense) than the original ones, affecting thus the overall typology of a language. The classical instantiation of this pervasive tendency is the diachronic strive to re-establish the One-Meaning–One-Form principle in “disturbing” situations where it does not obtain, e.g. in the presence of synchronically “superfluous” morphophonological alternations in inflectional paradigms. Such complexities are likely to arise in situations where the influence of L2 contact is weaker – situations often referred to as “normal” linguistic change, cf. Dahl’s (2004) concept of maturation.

It is however, not a trivial question how this kind of complexity related phenomena could be captured and described in a systematic fashion. What does complexity mean in the first place? Can it be objectively measured at all? Is the old hypothesis true that, overall, all languages are equally complex, and that complexity in one grammatical domain tends to be compensated by simplicity in another? These fundamental questions are addressed by many contributors to this volume.

Somewhat surprisingly, Nicholas Rescher's (1998) seminal philosophical analysis of what complexity is has never come to the fore in linguistic discussions of the issue. As a starter for conceptual clarification, we quote Rescher's (1998: 1) definition: "Complexity is first and foremost a matter of the number and variety of an item's constituent elements and of the elaborateness of their interrelational structure, be it organizational or operational." More specifically, Rescher (ibid. 9) breaks up the general notion of complexity into the following "modes of complexity" (our account is adapted to linguistically relevant matters and somewhat simplified):

1. *Epistemic modes*

A. *Formulaic complexity*

- a. *Descriptive complexity*: length of the account that must be given to provide an adequate description of the system at issue.
- b. *Generative complexity*: length of the set of instructions that must be given to provide a recipe for producing the system at issue.
- c. *Computational complexity*: amount of time and effort involved in resolving a problem.

2. *Ontological modes*

A. *Compositional complexity*

- a. *Constitutional complexity*: number of constituent elements (such as phonemes, inflectional morphemes, derivational morphemes, lexemes).

b. *Taxonomic complexity* (or *heterogeneity*): variety of constituent elements, i.e. number of different *kinds* of components (variety of phoneme types, secondary articulations, parts of speech, tense-mood-aspect categories, phrase types etc.).

B. *Structural complexity*

a. *Organizational complexity*: variety of ways of arranging components in different modes of interrelationship (e.g. variety of premodification or postmodification alternatives in basic constituent types such as noun phrases; variety of distinctive word order patterns).

b. *Hierarchical complexity*: elaborateness of subordination relationships in the modes of inclusion and subsumption (e.g. variety of successive levels of embedding and modification in phrases, clauses, and sentences; variety of intermediate levels in lexical-semantic hierarchies).

3. *Functional complexity*

A. *Operational complexity*: variety of modes of operation or types of functioning (e.g. variety of situational uses of expressions; variety of styles and speech situations; cost-related differences concerning language production and comprehension such as Hawkins' 2004 efficiency, etc.).

B. *Nomic complexity*: elaborateness and intricacy of the laws governing the phenomenon at issue (e.g. anatomical and neurological constraints on speech production; memory restrictions on sentence production and understanding).

Rescher's taxonomy provides a background for evaluating the coverage of the papers in this volume. Descriptive, constitutional, and taxonomic complexity all figure prominently in many of the contributions, especially under headings such as "length of description", "overspecification", and "absolute complexity". Organizational complexity surfaces in connection with "structural elaboration", in the form of redundant morphophonological alternations creating intraparadigmatic morphological complexity. Operational (i.e. processing-related) complexity is treated by some contributors under notions such as "relative complexity" or "efficiency". Hierarchical complexity, a central theme in formal linguistics (cf. the Chomsky hierarchy of languages, the most well-known linguistic measure of syntactic complexity), is not treated at any great length, reflecting the non-formalist approaches of the authors

2. An overview of the contributions in this volume

Three thematic areas emerged as focal from the ensemble of the papers, and these areas define the three sections into which the book is divided: "Typology and theory", "Contact and change", "Creoles and pidgins". These are naturally overlapping themes, but the placement of each paper in a given section is meant to reflect the central focus of the paper. The first

section, “Typology and theory”, unites eight papers concerned with typological comparison and with general theoretical issues such as the definition and measurement of complexity.

Continuing the discussion started in his 2003 doctoral dissertation, Wouter Kusters addresses fundamental issues in research on linguistic complexity. He argues for a relative approach to complexity, defining as complex those linguistic features that cause difficulties for L2 learners. He introduces the “general outsider” as an ideal L2 learner, neutralizing the effects of the mother tongue and cultural background, and proposes a set of concrete criteria for measuring complexity. The theoretical and methodological discussion is illustrated with a case study on the development of verbal inflection in selected Quechua varieties.

Matti Miestamo broaches a number of theoretical and methodological issues relevant for the cross-linguistic study of grammatical complexity. Two basic approaches to complexity are distinguished: the absolute one where complexity is seen as an objective property of the system, and the relative one: complexity as cost/difficulty to language users. The usability of these approaches in typological studies of complexity is evaluated. Miestamo argues that in typological studies of complexity it is better to focus on specific functional domains that are comparable across languages. Some general criteria for measuring complexity are introduced and evaluated, in particular the criteria of Fewer Distinctions and One-Meaning–One-Form. As for the relationship between complexity and cross-

linguistic rarity, it is likely that rarity shows at least some correlation with absolute complexity.

Gertraud Fenk-Oczlon and August Fenk approach complexity from a systemic perspective, continuing their earlier work on systemic typology (e.g. Fenk-Oczlon and Fenk 1999). They argue that languages vary in terms of complexity within subsystems but that trade-offs occur between the subsystems (phonology, morphology, syntax, and semantics). Analysing data on monosyllables they show that syllable complexity, number of syllable types, and the number of monosyllables correlate in statistically significant ways. Some aspects of semantic complexity are also discussed as well as the difficulty of talking about the complexity of rigid word order.

Kaius Sinnemäki scrutinizes the old idea that languages trade off complexity in one area with simplicity in another. He tests this hypothesis with a complexity metric based on the functional load of different coding strategies (especially head marking, dependent marking and word order) that interact in the marking of syntactic core arguments. Data from a balanced stratified sample of 50 languages show that the functional use of word order has a statistically significant inverse dependency with the presence of morphological marking, especially with dependent marking. Most other dependencies were far from statistical significance and in fact provide evidence against the trade-off claim, leading to its rejection as a general all-encompassing principle. Overall, languages seem to adhere more strongly to

distinctiveness (overt marking of important distinctions) than to economy (minimization of the use of overt markers).

Patrick Juola discusses several proposed definitions of complexity and shows how complexity can be assessed in these frameworks. As empirical testbed for developing a measure of complexity Juola uses translations of the Bible into English and 16 other languages. The variation in size of the uncompressed texts is substantially more than the variation in size after Ziv-Lempel (LZ) compression. This suggests that much of the variance in document size of the Bible is from the character encoding system. Juola argues that LZ, with its focus on the lexicon and long-term storage and retrieval of words, is a better model of the underlying regularities of language as expressed in corpora than linear complexity. Translations tend to be more complex (longer) than the original text. Furthermore, morphological and syntactic complexity can be measured by distorting the original text. Languages are about equally complex, but they express their complexity differently at different levels.

David Gil's paper treats the complexity of isolating vs. non-isolating languages. He presents an experimental approach to cross-linguistic analysis of complexity by investigating the complexity of compositional semantics in these language types. Compositional semantics is considered the simpler the more associational operations are available in a language. According to the compensation hypothesis, their availability should be roughly equal in the two language types but Gil's method reveals that isolating languages

generally allow for more associational operations than non-isolating languages. He further argues that the possible locus of complexity in isolating languages could not reside in pragmatics either, but that they have to be considered overall simpler than non-isolating languages, thus counteracting the claim that all languages are equally complex.

Elizabeth Riddle presents a diametrically opposed view on the complexity of isolating languages. Basing her discussion on data from three isolating languages in southeastern Asia, viz. Hmong, Mandarin Chinese and Thai, she argues that while these languages are indeed simple as regards bound morphology, they show considerably more complexity in lexical categories. Her concept of “lexical elaboration” encompasses phenomena such as classifiers, verb serialization, compounding, and a special type of elaborate expressions found in these languages. The discussion also addresses the fundamental issue of the borderline between lexicon and grammar.

Östen Dahl discusses complexity from the point of view of linguistic resources, that is, the set of possibilities the system offers its users. This aspect was left rather untouched in his (2004) book which laid out a framework for describing the increase of complexity in grammaticalization processes. The author argues that Sirionó, a Tupí-Guaraní language spoken in Bolivia, lacks syntactic NP coordination, wherefore its system offers its users fewer possibilities than many other languages do, and is thus less complex in this respect. He interprets the strategies employed by Sirionó as

incremental and those resembling the English *and*-strategy as compositional, arguing for a grammaticalization path and a complexity increase from the former to the latter.

The four papers in the section “Contact and change” place their focus on how language complexity is affected in situations of language contact and what happens to complexity when languages change.

John McWhorter continues the discussion started in his 2001 article in the journal *Linguistic Typology*. He has argued that the grammars of languages may undergo significant overall simplification only through large-scale non-native acquisition by adults. The extreme cases of such simplification are creoles, but less extreme simplification vis-à-vis their closest relatives can be observed in many non-creole languages used as lingua francas (e.g. English, Mandarin Chinese, Persian or Indonesian). In his contribution to the present volume, he discusses a few cases found in Indonesia that seem to present challenges to his views.

Casper De Groot argues that Hungarian spoken outside Hungary manifests an increase in complexity compared to Hungarian spoken in Hungary – typical of languages spoken in multilingual environments (cf. Nichols 1992). He shows that the general trend in the contact varieties is towards analyticity and periphrastic expressions and that the newly acquired structures are usually based on ones already present in the replica languages. De Groot adopts the methodology in Dahl (2004) arguing that these changes in Hungarian spoken outside Hungary manifest an overall decrease in

morphological complexity which, however, is accompanied by concomitant changes in syntactic complexity.

Eva Lindström's contribution is based on a total of around two years of fieldwork on the island of New Ireland in Papua-New Guinea. The author distinguishes between complexity as an objective property of the system, and difficulty experienced by adult learners of a language, and introduces explicit criteria for approaching both complexity and difficulty. She compares the Papuan language Kuot (her field language) to neighbouring Austronesian languages, discussing the interesting sociolinguistic situation where Kuot is being viewed as more difficult than its neighbours by the communities, and the use of Kuot is declining more rapidly than that of the neighbouring languages.

Antje Dammel and Sebastian Kürschner present an in-depth survey of morphological complexity of nominal plural marking in ten Germanic languages. They propose a metric that combines qualitative criteria, e.g. number of allomorphs, with more qualitative criteria, such as deviations from One-Meaning–One-Form, motivated e.g. by Natural Morphology. The combination of the multiple criteria shows the greatest accumulation of the most complex values in Faroese whereas the simplest values cluster the most in English. The authors further seek to validate the results by psycholinguistic experiments, a welcome appeal to bridge the so-called absolute and relative approaches to complexity (see Miestamo's and Kusters' papers in this volume).

The final section, “Creoles and pidgins”, contains four papers focusing on different issues of complexity and simplification in these contact languages, a central question being whether the contact past of these languages shows in their degree of complexity.

Mikael Parkvall addresses the complexity of creoles vs. non-creoles with the help of the database in the *World Atlas of Language Structures* (Haspelmath et al. (eds) 2005) – the largest typological database publically available. A method of quantification is introduced, and after selecting the suitable features, a large sample of languages is placed on a scale of complexity. Additional data on a number of pidgins and creoles is analysed and these languages are also placed on the scale. The author approaches his results from different viewpoints present in current creolistics and discusses their implications to the debate on creole origins. In addition to being the first large-scale attempt to quantify the typological profile of creoles vs. non-creoles, it is also a nice illustration of the unprecedented possibilities the WALS database offers for cross-linguistic research in different domains.

Harald Hammarström’s paper deals with the complexity of numeral systems and includes a case study of the domain in pidgins and creoles. This domain may be studied in a very broad cross-linguistic perspective since its semantics are exceptionally clear-cut and data is available from a very large number of languages – the array of languages considered in the paper is impressive even for a typologist reader. Hammarström examines the complexity of numeral systems, defining complexity in terms of minimal

description length. The complexity of pidgin and creole numeral systems is compared to their respective lexifiers and to the estimated cross-linguistic average. The paper is interesting in showing how a specific domain may be studied in terms of complexity, and it also connects to the on-going debate on the complexity of creoles vs. non-creoles.

Angela Bartens and Niclas Sandström apply the morphosyntactic 4-M model developed by Carol Myers-Scotton and Janice Jake to Cape Verdean Creole Portuguese (CVC). The 4-M model is claimed to be a universal model for the classification of morphemes according to their degree of cognitive activation in the process of utterance formation. CVC has been less strongly restructured than so-called prototypical creoles. The main topic is nominal plural marking, where CVC has morphosyntactic configurations similar to Brazilian Vernacular Portuguese (BVP). Previous accounts of CVC and BVP nominal plural marking mention the occurrence of (at least) one inflectional marker per noun phrase. Bartens and Sandström demonstrate that the reduction of inflectional plural marking in CVC results in massive loss of morphosyntactic complexity, due to CVC having arisen through substantial reduction and restructuring during creolization and to being young in comparison to its lexifier Portuguese.

Päivi Juvonen examines how the minimal lexicon of the pidgin language Chinook Jargon (CJ) gains maximal efficiency when used in a contemporary fictional text. CJ is well documented and even though its glory days with some 100,000 speakers in the 1880s are long gone, there are

still some speakers left in British Columbia and Northwest Oregon. A speaker of e.g. Swedish may have a passive vocabulary of some 60,000 words whereas many pidgin languages only have vocabularies comprising a few hundred words. The total number of lexical morphemes listed in different sources for such well documented pidgin languages as Chinook Jargon, Mobilian Jargon and Lingua Franca range from around 500 to just over 2000. The paper describes the CJ lexicon from a structural point of view and then examines the use of multifunctional lexical items in comparison to English. The results show, first, that there is no bound morphology (neither derivational nor inflectional) in the variety studied and, second, that there is much more multifunctionality in the pidgin text than in the English texts.

3. Acknowledgements

First and foremost, we would like to thank the authors for their interesting contributions and for the smooth and pleasant cooperation. Each paper has been read and commented on by at least two anonymous referees. We wish to express our gratitude to the colleagues who have devoted their time for refereeing the papers. We would also like to thank the editors of the Studies in Language Companion Series, Werner Abraham and Michael Noonan, for inviting us to publish this collection of articles in the series and for all their

work behind the scene. Thanks are due to Kees Vaes and the other people at Benjamins who have helped in making this book a reality. The Linguistic Association of Finland deserves special thanks for cooperation in organizing the conference that gave birth to the volume. We would also like to thank the conference participants for the lively and sophisticated scholarly exchange and the nice atmosphere they managed to create with their presence at the conference. For financial and institutional support, we would like to thank the Academy of Finland and the University of Helsinki.

References

- Chipere, N. 2003. *Understanding Complex Sentences. Native Speaker Variation in Syntactic Competence*. Basingstoke/New York: Palgrave Macmillan.
- Dahl, Ö. 2004. *The Growth and Maintenance of Linguistic Complexity*. Amsterdam: Benjamins.
- Deutscher, G. 2000. *Syntactic Change in Akkadian. The Evolution of Sentential Complementation*. Oxford: Oxford University Press.
- Everett, D.L. 2005. Cultural constraints on grammar and cognition in Pirahã. Another look at the design features of human language. *Current Anthropology* 46: 621-646.

- Fenk-Oczlon, G. and Fenk, A. 1999. Cognition, quantitative linguistics, and systemic typology. *Linguistic Typology* 3: 151-177.
- Haspelmath, M. 2006. Against markedness (and what to replace it with). *Journal of Linguistics* 42: 25-70.
- Haspelmath, M., Dryer, M, Gil, D. and Comrie, B. (eds) 2005. *The World Atlas of Language Structures*. Oxford: Oxford University Press.
- Hawkins, J.A. 1994. *A Performance Theory of Order and Constituency*. Cambridge: Cambridge University Press.
- Hawkins, J.A. 2004. *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press.
- Karlsson, F. 2007a. Constraints on multiple initial embedding of clauses. *International Journal of Corpus Linguistics* 12 (1): 107-118.
- Karlsson, F. 2007b. Constraints on multiple center-embedding of clauses. *Journal of Linguistics* 43 (2): XXX-XXX.
- Kusters, W. 2003. *Linguistic Complexity: The Influence of Social Change on Verbal Inflection*. PhD Dissertation, University of Leiden. Utrecht: LOT.
- McWhorter, J. 1998. Identifying the creole prototype: Vindicating a typological class. *Language* 74: 788-818.
- McWhorter, J. 2001. The world's simplest grammars are creole grammars. *Linguistic Typology* 5 (2/3): 125-166.
- Nichols, J. 1992. *Linguistic Diversity in Space and Time*. Chicago: The University of Chicago Press.

Perkins, R.D. 1992. *Deixis, Grammar, and Culture*. Amsterdam: John Benjamins.

Rescher, N. 1998. *Complexity. A Philosophical Overview*. New Brunswick and London: Transaction Publishers.