On the Feasibility of Complexity Metrics

Matti Miestamo, University of Helsinki

1. Introduction

The attention of the typological community has recently been attracted to questions of grammatical complexity. Most notably, McWhorter (2001a) introduced a metric for measuring overall grammatical complexity of languages in order to show that the world's simplest grammars are creole grammars; the target article was followed by eleven commentaries dealing with the issue from the point of view of creoles and more generally, and the author's response to these (McWhorter 2001b). This paper[1] does not focus on the differences between creoles and non-creoles. I will approach the question in more general terms and discuss some issues that come up in developing metrics for studying grammatical complexity from a general typological perspective, i.e. on the basis of an explicitly defined typological sample of languages. Here I will not go into detail about how such a sample should be constructed, but it can be noted that since the study of grammatical complexity requires very detailed analysis of the sample languages, the size of the sample cannot be very extensive.[2]

Although he needs a metric of overall complexity to support his thesis concerning the simplicity of creole grammars, McWhorter (2001a: 133) questions the theoretical relevance and usefulness of constructing a metric that could be used for measuring the grammatical complexity of the world's languages more generally. I agree that ranking the world's languages in terms of grammatical complexity is not an important goal in itself. Nevertheless, the question whether and how a metric of grammatical complexity could be constructed is intimately linked with our understanding (and the very definition of) linguistic complexity; tackling the question will therefore certainly increase our understanding of the nature of complexity and provide us with better tools for approaching the issue.

---

[1] My warmest thanks to the following persons for their comments on earlier versions of this paper: Fred Karlsson, Kaius Sinnemäki, John McWhorter, Wouter Kusters, and Guy Deutscher. Thanks are also due to the participants of the Complexity Workshop at the FinEst Conference of Linguistics, Tallinn Pedagogical University, May 7, 2004, and to the audience at the 20th Scandinavian Conference of Linguistics, University of Helsinki, January 7, 2004.

[2] Note also that a typologist, working with grammars, cannot take into account all the variation that a language shows across speakers, genres etc., which would certainly be relevant for overall grammatical complexity, but must concentrate on what could be termed core grammar, i.e. the set of phenomena covered by (good) reference grammars. Obviously grammars of roughly the same size and type must be used for all sample languages.

2. Complexity

The received view among linguists is that all languages are equally complex in their grammars. Differences can be found in different areas of grammar, but it is assumed that complexity in one area is always compensated by simplicity in another. Hockett (1958) phrases the view as follows (for a more recent formulation, see e.g. Crystal [1987: 6]):

"Objective measurement is difficult, but impressionistically it would seem that the total grammatical complexity of any language, counting both morphology and syntax, is about the same as that of any other. This is not surprising, since all languages have about equally complex jobs to do, and what is not done morphologically has to be done syntactically. Fox, with a more complex morphology than English, thus ought to have a somewhat simpler syntax; and this is the case." (Hockett 1958: 180–181)

This view is of course not shared by all linguists (see e.g. Dixon 1997: 4), and recently it has been explicitly challenged by e.g. McWhorter (2001a) and Kusters (2003). There is no principled reason why all languages should be equal in their overall complexity or why complexity in one grammatical area should be compensated by simplicity in another; it is unclear what mechanism could cut complexity in one area as soon as another area has become more complex, i.e. what could be the factor responsible for equi-complexity, see Kusters (2003: 9–12) for a discussion.

To prevent any misunderstandings, I want to emphasize that no value judgements are involved in talking about grammatical complexity. No claims are made about the efficiency or the expressive power of languages either. If some languages can be shown to have simpler grammars than others, this does not mean that, as tools of communication, these languages are either more efficient in lacking complexities that might make communication slower and more cumbersome, or less expressive, let alone primitive, in lacking features necessary for communication.

A study of overall complexity of languages must take into account all aspects of language from phonology to grammar and lexicon, and beyond. When focusing on grammatical complexity, one disregards things that can, at least for methodological reasons, be defined as being outside grammar. This concerns the lexicon for example.[3] It should be clear that when excluding the lexicon and other extra-grammatical aspects of language, one is not talking about complexity of languages in a broad sense, but only of grammatical complexity.

---

[3] It is of course difficult to draw a sharp line between grammar and lexicon. Although for some purposes it is useful to make this distinction, this does not necessarily mean that there is an ontological/substantial difference between the two. Rather, they exist in a continuum; grammar refers to the more general and productive items and processes whereas lexicon refers to the more particular and idiosyncratic ones.

There are two alternative approaches to linguistic complexity, which I call ABSOLUTE and RELATIVE, i.e. relative to language users. In the absolute approach, the definition of complexity refers to the number of parts in a system, or, in information-theoretical terms, to the length of the description of a phenomenon (cf. Dahl 2004). In the relative approach, complexity is defined in terms of difficulty of processing or difficulty in language acquisition or learning. McWhorter's approach to overall grammatical complexity can be characterized as primarily absolute (see esp. 2001a: 134–135 and 2001b: 395–396), whereas Kusters' (2003) approach to complexity in verbal inflection is of the relative type. Dahl (2004) also approaches complexity from the absolute point of view, grounding his notion of complexity in the principles of information theory.

Some linguistic phenomena can be difficult for a certain group of language users while facilitating the task of another group. To take some examples, redundant agreement may cause difficulty to speakers and L2 learners, but can be more helpful for hearers and L1 learners, while fission (one meaning expressed by many forms syntagmatically) is preferred by hearers but causes difficulties for the three other types of language users (see Kusters 2003: 51–52, 56–57). Therefore, if one approaches complexity from the relative point of view, one has to ask the question "complex to whom?". Kusters (2003: 6–7) defines complexity in relation to adult learners of language, and those properties of languages that cause difficulties for L2 learners are therefore criterial for complexity. Kusters' complex-to-whom question is very important, and in his own work he is correct in focusing on L2 learners; in examining the simplifying effects of sociolinguistic factors such as language contacts the problems of L2 learners become the key issue. However, if we want to reach a definition of complexity that could be applied as generally as possible, the primary relevance of L2 learners is by no means obvious. In fact they could be considered the least important of the four groups mentioned — every natural language has (or at least has had) speakers, hearers, and L1 learners, but L2 learners are usually a much more marginal group. There will always be some conflict between definitions of complexity based on different types of users, and no general user-type-neutral definition is possible. This is problematic for a relative approach to overall complexity: Should one of these groups be chosen as criterial for the definition of complexity? How to motivate this choice?

When studying overall grammatical complexity from a cross-linguistic point of view, a relative approach to complexity faces some further problems. In principle, a cross-linguistic study of overall grammatical complexity has to be able to take into account all aspects of grammar found in all languages. This involves a vast array of phenomena, and there is no way that all aspects of all grammars could be characterized as to what kinds of difficulties they cause to different types of language users. In typological studies of overall complexity one cannot undertake actual acquisition studies or psycholinguistic experiments concerning the linguistic properties found in the languages under study. Using principles based on what is

known from previous studies of acquisition or processing (as e.g. Kusters does), some subset of the properties found in the languages under study can certainly be identified as causing difficulties to different types of language users. This subset will however not be sufficient for a typological study of overall complexity.

These problems suggest that if one wants to study overall complexity from a typological point of view, one should leave the complex-to-whom question aside and define complexity in absolute terms, paying attention to system size (or the length of its description), e.g. categorial elaboration, number of units, number of rules. Whether and how the different aspects of complexity defined this way contribute to ease vs. difficulty of language use for different types of users is a separate question to be addressed with the help of psycholinguists and researchers of language acquisition and learning; this will then also serve as one way of testing the theoretical and explanatory relevance of the absolute definition of complexity.

3. A Metric of Overall Grammatical Complexity: McWhorter (2001a)

Not many have tried to compare languages in terms of overall grammatical complexity. The most detailed and comprehensive attempt found in the literature is the metric proposed by McWhorter (2001a). The metric is designed for measuring the overall grammatical complexity of languages in order to support the thesis that the world's simplest grammars are creole grammars. It pays attention to overt signaling of phonetic, morphological, syntactic and semantic distinctions beyond communicative necessity: a grammar is more complex than another to the extent that 1. its phonemic inventory has more marked members, 2. its syntax requires the processing of more rules, 3. it gives overt and grammaticalized expression to more fine-grained semantic and/or pragmatic distinctions, and 4. to the extent it uses inflectional morphology. As stated in the introduction, the present paper is not the first discussion of McWhorter's ideas, and my main focus being on the general principles of studying grammatical complexity, I will not deal with all aspects of McWhorter's proposal in detail. I will only briefly comment on the inflectional criterion which is not as clear from the above formulation as the other three criteria are. McWhorter (2001a: 137–138, 2001b: 397) points out that inflection does not necessarily increase complexity; expression of grammatical categories by purely agglutinative inflections is not more complex than analytic expression of the same categories, but inflection often renders a grammar more complex in that it is likely accompanied by complexifying factors such as morphophonemics, suppletion, and allomorphy. A more explicitly formulated treatment of complexities in inflection can be found in Kusters (2003), where violation of any of the following three principles (and a few additional less central ones) is taken to increase complexity: 1. Economy — restriction of the number of overtly signalled categories, 2. Transparency — clarity of the relation between meaning and form (one-meaning-one-form), and 3. Isomorphy — identity of the order of

elements in different domains.[4] Note that although Kusters uses these principles in a relative approach to complexity, they are to a large extent compatible with (and to some extent also based on) an absolute definition of complexity; e.g. Economy is about the number of overtly signalled categories and thus deals with system size, and when Transparency is violated, the description of the relationship between meaning and form becomes longer.

McWhorter's metric is not intended as a fully developed device capable of determining the overall complexity of the grammar of any language in exact terms, but rather as a preliminary tool sufficient to show the clear complexity difference between the creole and non-creole languages in his study. As applied by McWhorter, the metric works rather well for the purpose it is designed for: two languages are compared at a time, one creole (Saramaccan) and one non-creole (Tsez, Lahu), and the creole appears less complex according to each criterion of the metric. When the differences are as clear as they are between the languages compared by McWhorter, it seems plausible to draw the conclusion that some grammars are more (or less) complex than others. When each criterion gives the same result, i.e. designates language A as more complex than language B, and there is no conflict between the criteria, then if the criteria can be thought to represent the overall grammatical characteristics of languages with sufficient coverage, language A can be said to be overall more complex in its grammar than language B.

In the following I will address the question whether similar principles can be used for purposes the metric was not originally developed for, i.e. whether a metric can be constructed for comparing the overall complexity of grammars from a general typological point of view. Just as when comparing creoles and non-creoles, a metric based on similar principles as those underlying McWhorter's proposal can in principle reveal clear complexity differences even when comparing a larger number of languages whether creoles or non-creoles: if the grammar of one language is systematically more (or less) complex than those of other languages according to each criterion of the metric, one can draw the conclusion that that language has a more (or less) complex grammar overall compared to the other languages. But when overall complexity differences are more subtle, and the criteria give conflicting results, the issue becomes more problematic. The usefulness of such a metric naturally hinges on the question how often complexity differences between languages are so clear that the metric can reveal them. More subtle differences are to be expected when comparing a larger number of languages, not limiting comparison to differences between creoles and non-creoles.

---

[4] In terms of these three principles, Kusters examines several aspects of verbal inflection susceptible of causing difficulty to different types of language users and that could thus be potentially usable as criteria for complexity: redundant agreement, non-redundant agreement, expression of tense/aspect/mood, expression of voice (Economy); morphological allomorphy, accidental homonymy, fission, fusion, phonological allomorphy, structural homonymy (Transparency); marked affix order, inconsistent affix order (Isomorphy).

4. Problems in Studying Overall Grammatical Complexity

In this section I will discuss two general problems that any metric of overall grammatical complexity has to face. The first one, which I will call the PROBLEM OF REPRESENTATIVITY, is quite obvious. To measure overall grammatical complexity, the criteria in a metric should take into account all aspects of grammar as exhaustively and in as much detail as possible. Obviously, this is beyond the capabilities of current linguistic theories, and even if it became theoretically possible to somehow characterize every aspect of a grammar of a language, this would be too vast a task for any linguist to accomplish, and even more so when dealing with a typological language sample. From a more practical point of view, one can ask how closely one should be able to meet this ideal in order to be able to say something about overall complexity. One may arrive at a level of representativity sufficient to reveal clear complexity differences between grammars, but the more fine-grained distinctions one is dealing with, the more problematic the issue of representativity becomes. The problem of representativity has also been identified by McWhorter (2001a: 134), where he notes that the level of detail in his criteria is not sufficient to reveal finer complexity differences than those found between the creoles and non-creoles that he compares. It can also be noted that in a typological study the representativity issue is linked with the comprehensiveness of reference grammars (cf. footnote 1).

The problem of representativity causes severe difficulties for metrics of overall grammatical complexity, but it may be resolvable in the sense that a satisfactory level of representativity can possibly be reached. The second problem, which I will call the PROBLEM OF COMPARABILITY, is more serious and may well be irresolvable. As discussed above, when applying a metric to a typological language sample, we expect the complexity differences between most of the languages to be less clear than between the languages compared by McWhorter, and the different criteria used in a metric are then likely to give conflicting results. We are thus faced with the question of the relative weight of the criteria used, i.e. how much each criterion contributes to the overall picture. If we want to ask the question of overall grammatical complexity and see whether complexity in one area is compensated by simplicity in another, the criteria should be explicitly quantified in such a way that the numbers be mutually comparable — the relative weight of each criterion should be measurable, and it should be possible to add up these numbers to get an overall complexity value for each language. The problem is not quantifying the criteria — McWhorter's criteria are inherently quantitative: he talks about the number of marked phonemes, of rules and of semantic and pragmatic distinctions, as well as about the extent to which a language uses inflection — but the problem is how to make the numbers mutually comparable. Is such comparability possible, and what would the commensurability of the numbers be based on? These are crucial questions for any attempt to measure overall complexity.

A metric using criteria such as the ones proposed by McWhorter (2001a), for example, must address the question how phonological or morphological complexity compares with syntactic complexity or with the number of grammaticalized semantico-pragmatic distinctions. Would a language with a lot of complexity in its inflectional system but a more restricted number of grammaticalized semantico-pragmatic distinctions be equal in overall complexity with a language where the opposite pattern was found? Similar questions arise inside the areas that the criteria refer to, not only between them. In connection with the criterion concerning the number of grammaticalized semantico-pragmatic distinctions, for example, one has to address the question how the numbers of distinctions made in different functional domains can be compared to each other; a language that makes five aspectual distinctions is more complex in this respect than one that makes only two, but languages grammaticalize different functional domains, some tense, some aspect, some evidentiality, some make many distance distinctions in their deictic systems, and these different domains are not commensurable — how should we compare the number of aspectual distinctions to the number of distinctions made in the deictic system, how many tense distinctions would be equal to how many aspectual ones? The problem of comparability is not addressed by McWhorter (2001a) or his commentators, although Arends (2001: 181) touches upon it when discussing "the issue of the relative weight of the number of rules as opposed to their internal complexity" in connection with McWhorter's syntactic criterion.

The problem of comparability does not invalidate the point McWhorter wishes to make in his paper, because the complexity differences between the languages compared are so clear; there is no need to decide how much each criterion weighs because the criteria do not give conflicting results. With each pair of languages, one language is always more complex than the other according to each criterion, and thus overall clearly more complex according to the metric. But the problem cannot be escaped when complexity differences between languages are smaller. Language A can be more complex than language B according to criteria 1 and 3, while language B is more complex according to criteria 2 and 4. The results given by the different criteria will show such discrepancies when the complexity differences between the languages examined are not as clear as in the cases examined by McWhorter. The incommensurability of the different areas of grammar will make all attempts at fine-grained measurements of overall complexity wellnigh impossible.[5]

---

[5]

Comparability could in principle be achieved using a relative definition of complexity if one knew everything about the grammar of each sample language, and everything about what happens in the human brain when language is processed, encoded or decoded, or when language is being acquired. A measure of the costs of the totality of these processes could give us a tool to measure and compare overall grammatical complexity. While we are waiting for the scientific community to arrive at this knowledge, the cross-linguistic study of overall complexity has to rely on an absolute definition of complexity, using a metric such as the one proposed by McWhorter and contenting itself with the fact that only clear differences in overall complexity can be revealed by such metrics.

5. Conclusions and Future Prospects

Any metric of overall grammatical complexity is subject to the problems of comparability and representativity — how to ensure that the criteria used are mutually comparable and that they account for all aspects of grammars exhaustively. The problem of representativity alone makes it extremely difficult to make more fine-grained characterizations of overall grammatical complexity. The problem of comparability has even more serious consequences for the methodology of studying overall complexity. We can only compare what is comparable, and because different areas of grammar are not comparable with each other in terms of their contribution to overall complexity, we must focus on the complexity of specific areas of grammar instead of overall complexity, limiting cross-linguistic comparison to these specific areas. The cross-linguistic study of complexity is therefore necessarily primarily concerned with the complexity of specific areas of grammar, and the question of overall complexity is secondary.

To conclude this paper, I will make some more specific suggestions as to how the typological study of grammatical complexity could proceed. As discussed above, the puzzle of overall grammatical complexity has many pieces, and whether or not one aims at resolving the whole puzzle, one has to work piece by piece. Grammatical structure can be divided into different areas in many different ways. I propose the following principles for subdividing the areas of grammar. One can first make a division along the lines of the double articulation of language, i.e. approach phonology and morphosyntax separately. I will not say more about the complexity of phonological systems and processes in this paper. In morphosyntax the study of grammatical complexity can be approached from the point of view of functional domains. In typological studies cross-linguistic comparability is usually guaranteed by basing comparison on function, not form, i.e. the formal encoding of similar functions, such as negation, can be compared across languages (see e.g. Stassen 1985: 14–15). This approach can also be used when studying complexity from a typological point of view. The first step is therefore to identify the functional domains encoded by each language in the sample. The functional domains can then be studied from two points of view: a) the study of grammatical meaning (cf. Frawley 1992), and b) the study of the formal encoding of grammatical meaning. Comparisons can then be made between similar functional domains across languages.

The study of grammatical meaning consists first and foremost of identifying what semantico-pragmatic functions are grammatically encoded, i.e. grammaticalized, in the sample languages in each functional domain, and criteria of complexity include the number of grammatical distinctions made in each functional domain and the cross-linguistic markedness of the encoded categories (cf. McWhorter's criterion 3 and Kusters' Economy). The study of the formal encoding of grammatical meaning deals with the relationships between the grammatical meanings encoded in a language and the forms that encode these meanings, i.e.

one studies the complexity of the formal morphosyntactic encoding of the functions grammaticalized in each functional domain. Criteria of complexity include the transparency of the formal encoding of the functions, patterns of irregularity (cf. McWhorter's criteria 2 and 4, and Kusters' Transparency), and identity of the order of elements in different domains (cf. Kusters' Isomorphy). In connection with both grammatical meaning and the formal encoding of grammatical meaning, it is also important to study how the functional domains interact with each other. Attention can be paid to asymmetries between functional domains (see Miestamo 2003), e.g. whether some distinctions made within a functional domain are neutralized in connection with another functional domain. Distinctions made in the functional domain of tense or aspect can for example be neutralized in connection with the functional domain of polarity (i.e. under negation). This also helps us to shed more light on the question whether (and how) complexity in one domain is compensated by simplicity in another.

The cross-linguistic comparison of similar functional domains according to these criteria allows us to draw cross-linguistic conclusions about the complexity of the domains. As languages can grammaticalize very different functional domains, comparability is not always easy to arrive at. One can only compare functional domains that can be identified as coding essentially the same functions. Characterizing the cross-linguistic variety in the complexity of each functional domain and the interactions between domains are the primary concerns of the cross-linguistic study of complexity. Trying to combine the results from the different functional domains in order to be able to characterize overall grammatical complexity of languages faces the problem of comparability. The contributions of the different domains to overall complexity are difficult to define. But if there are very clear differences in overall complexity, these can be shown by such comparisons: if one has examined a significant number of functional domains in the sample languages in sufficient detail, and if one language turns out to be more (or less) complex than the others in all (or nearly all) of the functional domains examined, then we are perhaps entitled to draw the conclusion that the grammar of this language is more (or less) complex overall than those of the others.

References

Arends, Jacques 2001. Simple Grammars, Complex Languages. – Linguistic Typology 5, 180–182.

Crystal, David 1987. The Cambridge Encyclopedia of Language. Cambridge: CUP.

Dahl, Östen 2004. The Growth and Maintenance of Linguistic Complexity. Studies in Language Companion Series 71. Amsterdam: Benjamins.

Dixon, Robert M. W. 1997. The Rise and Fall of Languages. Cambridge: CUP.

Frawley, William 1992. Linguistic Semantics. Hillsdale, NJ: Lawrence Erlbaum Associates.

Hockett, Charles F. 1958. A Course in Modern Linguistics. New York: Macmillan.

Kusters, Wouter 2003. Linguistic Complexity, the Influence of Social Change on Verbal Inflection. Utrecht: LOT.

McWhorter, John H. 2001a. The World's Simplest Grammars are Creole Grammars. – Linguistic Typology 5, 125–166.

McWhorter, John H. 2001b. What People ask David Gil and Why: Rejoinder to the Replies. – Linguistic Typology 5, 388–412.

Miestamo, Matti 2003. Clausal Negation: A Typological Study. Ph.D. Dissertation, University of Helsinki.

Stassen, Leon 1985. Comparison and Universal Grammar. Oxford: Blackwell.