Linguistic complexity: interfaces and processing

The study of language has always centered around two sets of observable facts: the ease with which children acquire their native languages despite the wide typological variations that one can observe between these languages (e.g., French, Mohawk, Greenlandic, Gungbe, Kikongo, Xhosa, Mandarin Chinese, etc.), and the difficulties late learners (namely adult L2 learners) face in learning these same languages. Set against the difficulties that linguists have in establishing the (number of) rules necessary to describe some phenomena cross-linguistically, these two sets of facts about acquisition appear to relate to another research question that has led to a debate in the field for the past fifty years or so: How to account for the complexity of linguistic systems and to what extent do languages differ in complexity? Though different linguistic approaches exist that account for typological variation and related issues of acquisition, an implicit consensus in the field has been to assume that all languages are equally complex even though they may show different degrees of complexity in different modules (i.e., morphology, phonology, syntax, semantics, and pragmatics). Charles Hockett (1958: 180–1), cited in Sampson (2009: 2), expressed the consensus as follows:

> 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.

Recently, various views have been proposed in the literature which challenge this consensus, thus suggesting that languages can differ fundamentally in complexity and that the degree of complexity of a language may correlate with the context in which it emerged (e.g., L1 acquisition versus L2 acquisition), its age (e.g., pidgins/creoles versus older languages), or the size and social structure of its speaking community (small tightly related communities versus large and loose communities), see e.g., Bickerton (1981, 1984, 1988), Dahl (2004), Trudgill (2011), among many others. In such views, languages acquire more complex systems as they evolve, hence notions such as growth or maturation (or grammaticalization) as argued for in Dahl (2004: 2).

> In this book, I look at grammaticalization in the perspective of what I call maturity—mature linguistic phenomena being those that presuppose a non-trivial prehistory: that is, they can only exist in a language which has passed through specific earlier stages. Grammatical maturation—processes that give rise to phenomena that are mature in this sense—in general adds to the complexity of a language […] Complexity is here seen, not as synonymous with "difficulty" but as an objective property of a system—a measure of the amount of information needed to describe or reconstruct it.

Applied to the following examples from English and Gungbe, we may reach the conclusion that the morphosyntax of the English past verb is more complex than that of the Gungbe one. In English, one needs to specify that regular verbs must take an additional affix to encode past, while no such affix is required in Gungbe: the verb is always bare in this language and past is the default interpretation for eventive (or so-called dynamic) verbs (cf. Aboh, 2004a).
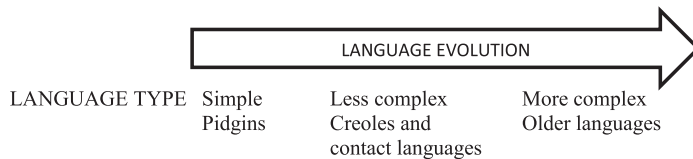
(1) a. John cook-ed   beans

      b. Ján ɖà   àbɔ̀bɔ̀

While this characterization of complexity may sometimes suggest that more superficial distinctions in a language may increase complexity, this need not be the case since what matters here is the amount of information required to describe or reconstruct the system (cf. Dahl, 2004, Audring, this volume).

1. Against a surface approach to complexity

Yet, common views of complexity tend to primarily focus on 'visible' (or audible) cues that the linguist can incorporate in her complexity metrics. Such views therefore take a surface approach to complexity. As is often the case, surface cues mainly involve morphological distinctions and related morphosyntactic rules. In this regard, a set of languages that are often cited in the literature as being systematically less complex because having less morphosyntactic, phonological and semantic distinction are creole languages and similar nativized contact languages (but see Dahl, 2004, chapter 6, Aboh and DeGraff, 2016 for a critique). Such ideas about creole simplicity derive from the much entertained, though lacking supportive evidence, pidgin-to-creole cycle hypothesis in creolistics: creoles being nativized vernaculars have more distinctive properties than their pidgin predecessors. Consequently, the description/reconstruction of creoles would require more information than needed for describing pidgins. Because of a shallow history, which does not give them enough time to acquire mature phenomena, creoles and similar contact languages would in turn require less information for their description/reconstruction than one would need for older languages (cf. McWhorter, 2001, 2011). According to this scenario, the complexity cline in human languages can therefore be represented as in (2).

(2)

LANGUAGE TYPE   Simple   Less complex   More complex
                 Pidgins   Creoles and   Older languages
                           contact languages

For scholars like Derek Bickerton, this developmental path makes pidgins and creoles the most relevant empirical domain for studying the emergence of language in human species (but see Mufwene, 2001 for a critique). Since complexity appears to grow in human language, this view implies that it is an objective property that linguistic systems wear on their sleeves which the linguist can list and evaluate against her complexity metrics. As a consequence, it is not uncommon that complexity metrics result in a list of overt morphosyntactic manifestations that one can detect in a language as compared to another. An example of such a surface complexity metric can be found in McWhorter (2001: 128) who after comparing the expression of past tense in Kikongo and Japanese claims that "Kikongo, in happening to have evolved as fine-grained an overt subdivision of pastness […] has a more complex past-marking system than Japanese." Applied to grammar as a whole, this view led McWhorter (2001: 135, 136) to the following metrics:

First, a phonemic inventory is more complex to the extent that it has more marked members.

Second, a syntax is more complex than another to the extent that it requires the processing of more rules, such as asymmetries between matrix and subordinate clauses (e.g., Germanic verb-second rules), or containing two kinds of alignment rather than one (i.e., ergative/absolute and nominative/accusative).

Third, a grammar is more complex than another to the extent that it gives overt and grammaticalized expression to more fine-grained semantic and/or pragmatic distinctions than another.

Fourth, inflectional morphology renders a grammar more complex than another one in most cases. […] Inflection more often than not has wider repercussions in a grammar […] which are complexifying factors in terms of exerting a load upon processing.

Under these claims, the size of a paradigm (e.g., morphology) or the number of putative morphosyntactic rule combinations and their related morphological exponents (e.g., verb placement and verbal inflection) is sufficient to evaluate linguistic complexity because these are "complexifying factors in terms of exerting a load upon processing". Though processing is called upon in defining some of his complexity metrics, McWhorter provides no experimental data to support his claims. Instead, much of his discussion focuses on distinctions that can be detected on surface forms. This surface approach to linguistic complexity is not unproblematic though.

2. Beware the internal/external syntax

First a surface approach to complexity mixes levels of analysis thus blurring a distinction made in Dahl (2004) and relevant studies between 'objective complexity' and 'difficulty'. Likewise, many chapters in this volume show that the claim "more morphology (or surface distinctions) implies more processing" must first be demonstrated on an empirical ground, and the correlation does not systematically hold for all relevant cases. On conceptual ground, such a list of features as provided by McWhorter and related studies implies that a pair like (3) involves two sentences of a similar structure. In French and English, the verbs are marked for third person singular in present and the two sentences contain the same number of arguments and modifiers:

(3) a. Jean mange du   pain   dans son   bureau

     b. John eats   bread in   his   office