1 Complexity in Language: A Multifaceted Phenomenon

Salikoko S. Mufwene, Christophe Coupé, and François Pellegrino

1 Complexity in Linguistics

1.1 Linguistics and the Science of Complexity

Complexity has attracted a great deal of attention in linguistics since 2001, at a rate that proportionally far exceeds its invocations in the field since Ferdinand de Saussure, the father of our discipline, in the early twentieth century. The number of books bearing *complexity* in their title is remarkable, suggesting that there may be an emergent research area whose focus is COMPLEXITY in Language. The dominant question that the relevant linguists have addressed is the following: To what extent does complexity as observed in different languages or in different modules of the language architecture display both cross-systemic variation and universal principles? This has entailed asking whether there are languages that are more complex than others and explaining the nature of differences.

One is struck by the sheer number of book-length publications alone,[1] and even more when the numerous journal articles and chapters in edited volumes are added to the total count, regardless of whether or not they include *complex(ity)* in their title. On the other hand, one is also shocked by the scarcity of works that explain what COMPLEXITY is, apparently because it is assumed to be known.[2] This is quite at variance with publications outside linguistics, which are devoted to explaining various ways in which the notion can be interpreted.

[1] Book titles containing the term *language* or *linguistic(s)* include: Dahl (2004), Hawkins (2005), Risager (2006), Larsen-Freeman (2008), Miestamo et al. (eds., 2008), Sampson et al. (eds., 2009), Givón (2009), Givón & Shibatani, (eds., 2009), Pellegrino et al. (eds., 2009), Aboh & Smith (eds., 2009), Faraclas & Klein (eds. 2009), Cyran (2010), Trudgill (2011), Robinson (2011), McWhorter (2012), Kortmann & Szmrecsanyi (eds., 2012), Housen & Kuiken (2012), Blommaert (2013), Culicover (2013), Massip-Bonet & Bastardas-Boada (eds., 2012), Newmeyer & Preston (eds., 2014), Berlage (2014), Kretzchmar (2015), and Baerman, Brown & Corbett (2015).

[2] A noteworthy exception is Ellis and Larsen-Freeman's (2009) "*Language as a complex adaptive system*," which is derived from the lead and seminal chapter by Beckner et al. (also identified as "The Five Graces Group"), "Language is a complex adaptive system: A position paper."

The subject matter has actually also evolved into what is identified by some as "complexity theory" or the "science of complexity" (see footnote citations).

Thus, a convenient starting point for this chapter and this book is to explain what is meant by *complexity* as it applies both to linguistics and other research areas. Etymologically, the term *complexity*, as a nominalization from *complex*, can ultimately be traced to Latin *complexus*, a past participle of the deponent verb *complecti* 'embrace, comprise,' according to *Webster's Collegiate Dictionary*, and also confirmed by the *French Petit Robert*, which translates it as *contenir* 'contain.' According to the *Online Etymology Dictionary*, the adjective *complex* 'composed of parts' was borrowed from French *complexe* 'complicated, complex, intricate' (seventeenth century), from Latin *complexus* 'surrounding, encompassing,' past participle of *complecti* 'to encircle, embrace.' In transferred use, the verb meant 'to hold fast, master, comprehend', from *com-* 'with' and *plectere* 'to weave, braid, twine, entwine.' The noun *complex* evolved to mean 'a whole comprised of parts.'

This etymological definition remains very generic. Beyond it, it appears that no strong consensus has emerged in the science of complexity itself about what *complexity* means (see, e.g., Strogatz 2003; Gershenson, ed. 2008; Mitchell 2009). There are nonetheless some common themes and properties that recur in the relevant literature. They include the following, which overlap in some ways:

(1) Complexity arises from the coexistence of components that interact with each other, not necessarily from the fact that a space or a system is populated with several components or members; it is therefore interactional.
(2) Complexity arises from the dynamics of activity coordination or synchronization that integrate individuals as members of a population (e.g., ant colonies, bird flocks, and fish schools); thus, it is dynamical.
(3) Complexity emerges from nonlinear evolution, which is driven by multiple factors whose significance may vary at different stages of the evolutionary process; its effects are not constant, but subject to the changing values of the relevant variables.
(4) Complexity lies in what brings order out of chaos,[3] through what is also known as "self-organization" and was formerly referred to as an "invisible hand" (Smith, 1776).[4]
(5) There is complexity in any system where the properties of the whole do not amount to the sum of the properties of the components.

[3] "Chaos" is used here as in "chaos theory," which studies systems whose behavior is highly sensitive to initial conditions, in the sense that small differences in initial conditions may produce quite divergent evolutions or outcomes. It also seeks to capture emergent patterns from the interplay between order and disorder, from which complexity arises.
[4] For a more linguistic take on the "invisible hand," see Keller (1994).

(6) Finally, complexity is the peculiarity of emergent patterns in a system in constant state of flux between disorder and transient order (or equilibrium). In other words, complexity arises from the dynamics of coexistence and interaction or cooperation of components toward generating the properties of whole.

As we are reminded by Loureiro-Porto and San Miguel (Chapter 8), *complex* should not be confused with *complicated* (*pace* the etymology of *complex(ity)* cited earlier). For instance, airplanes are complicated rather than complex pieces of engineering. Despite the very large number of parts, each part has a clear function that makes it possible – and to some extent easier – to predict its contribution to the whole. On the contrary, it is not evident which role each component of a true complex system (such as an ant colony or a flock of birds) plays in the behavior or function of the whole system, or what it specifically contributes as a unit to the larger, integrated whole.

Connecting these interpretations to Language, the idea that a system consists of interacting components is not new. It was indeed at the core of the structuralist program, in which phonemes, words, and other linguistic units were primarily considered as components of structures. An important peculiarity of this tradition is that a language was typically construed as an autonomous system, independent of its speakers and the wider ecology in which it and the speakers evolve. Thus, internal forces and their interactions were paid much more attention than external ones. Although this suits the etymology of the term *complexity* and some of the earlier points, it understandably omits other interpretations made evident in the science of complexity, especially regarding the dynamical aspects. Complexity arises not just from how the different parts interact with each other but also from how they respond to external pressures of the environment, or the external ecology (Mufwene 2001), outside the system.

However, decades later, despite the increasingly interdisciplinary nature of the relevant scholarship in other research areas, most linguists deal with complexity almost as if hypotheses in those other areas couldn't possibly apply to languages. This observation does not include modelers, often coming from the field of artificial intelligence, who have invoked multi-agent systems or network theory to investigate language emergence and change (e.g., de Boer & Zuidema (2010), Ke et al. (2008), Kirby (2000), Steels (1998, 2011a). Aside from them, others such as Massip-Bonet and Bastardas-Boadas, eds. (2012) or Kretzschmar (2015) have also highlighted the dynamic aspects of language behavior. Overall, linguists still have to ask themselves what interpretation of LANGUAGE they subscribe to, what COMPLEXITY may mean under that particular interpretation, whether the architecture of Language and linguistic behavior are exceptional in relation to the common properties of complexity that have been observed in other aspects of nature, and how they contribute to this expanding, more inclusive research endeavor.

1.2 Components, Structures, and Domains in Linguistic Complexity

Likely because most of them have ignored work in the science of complexity, linguists have remained faithful to the interpretation of a COMPLEX SYSTEM as 'a whole consisting of several parts.' Thus, the more parts the whole consists of, the more complex it is assumed to be, regardless of how the parts interact with each other. This explains why an approach commonly found in the literature consists in counting the number of elements of a linguistic system in order to evaluate its complexity. Depending on the specific study, the focus may be the number of phonemes, morphemes, or words, but also relations among variants of such units (allophones, allomorphs, or near-synonyms), or yet the number of categories, rules, or constraints that can be posited in a system. This has come to be called "bit complexity" and has been criticized as uninformative (DeGraff 2001, 2009).

Indeed, this approach does not pay attention to possible relationships between the components, and goes counter to the well-known idea that "simply more does not mean more complex." For example, a set of five bodies moving randomly – in the absence of any interaction force – is not as complex as, let alone more than, a set of five bodies, or even three or four, moving according to gravitational forces they exert on each other (Poincaré 1891).

The possibility that a whole with fewer parts engaged is several multilateral interactions can generate more interactive complexity cannot be accounted for in a bit-complexity approach. A good example is when an item generates different interpretations depending on what other item it is combined with. This is illustrated with the particle *up* in combinations with various verbs such as in *pick up*, *give up*, *show up*, and *look up*. While the item *up* is basically the same particle in all these constructions, its contribution to the meaning of each phrase appears to vary. This variation suggests that the particular dynamics of each combination produce the meaning of the whole phrase. The overall meaning of each phrase is not the sum of the meanings of its parts (see, for instance, Victorri 1994 on the dynamics of such constructions).

To be sure, some linguists have shifted from counting elements to assessing how they make a system together. These linguists have first attempted to identify the patterns of interactions between the components and then infer linguistic complexity from the interactions. Their approach has involved building mathematical graphs and then quantifying their "structural complexity" with an appropriate measure. The vertices of the graph correspond to the components of the system, while the edges connecting them reflect how they can be related meaningfully. Another strategy is presented in Coupé et al. (Chapter 6), in which patterns of co-occurrences of phonetic segments are evaluated with respect to the individual occurrences of these segments, in order to detect significant interactions.

Rather than focusing on (mostly pairwise) interactions between elements, another systemic view seeks to describe the linguistic system in terms of regularities and irregularities. A classical concept here is *Kolmogorov complexity*, which is the length of the shortest program that can produce the description of the system (given a programming language). Behind it is the central idea that the more compressible a piece of information is, the lower its complexity is as well. This algorithmic approach to complexity (Dahl 2004:42) is more processual than the previous one, because the algorithm has to be run in order to get the description. While it has been proved that the Kolmogorov complexity cannot be computed, reasonable approximations can be obtained with standard compression algorithms and be applied to compare complexity between objects. The size of an archive containing the compressed version of the initial description of the system and the means to decompress it (i.e., a self-extracting archive) is inversely proportional to the complexity of the system. Rissanen's (1978) "minimum description length" is another possible approximation to Kolmogorov complexity.

Such approaches have been applied especially to measuring the complexity of morphological systems (e.g., Bane 2008, Walther & Sagot 2011). Some linguists have also echoed Gell-Mann's (2003) concern that Kolmogorov complexity is highest in the case of totally random expressions, while we intuitively do not see totally unordered systems as complex. His counterproposal for measuring complexity effectively, named *effective complexity*, is the length of the shortest description of the set of regularities of the system. Along these lines, Newmeyer and Preston (2004:182) also state that "the more patterns a linguistic entity contains, the longer its description, and then the greater its complexity."

Although these quantitative approaches offer more refined considerations of linguistic complexity, they rely on the descriptions that linguists can provide of a linguistic system. When different options compete in this regard, the quantification methodologies themselves do not help. This echoes Edmonds' (1999) statement that complexity lies before all in the eye of the interpreter of the system. Another way of considering this is that a descriptive account of complexity can be at odds with a more functional approach: Does the complexity of the description of an utterance always correspond to the amount of difficulty the hearer experiences in processing it? Does the description capture adequately the nature of the neural and psychological encoding, and its consequences in terms of processing? Is such an approach to complexity informative about the overall complexity of a language?

Attempts to assess the complexity of a whole language present an additional difficulty with the same endeavor in other, physical or cultural systems. For example, ferromagnetic materials, a well-known physical system in which self-organization of microscopic magnets can occur in the absence of strong external magnetic field, are only composed of identical elements without hierarchical structure. By contrast, in the case of languages, different modules and levels of analysis – phonology, morphology, syntax, semantics, and so on – can be distinguished, and the hierarchical integration of their elements (starting from meaningless sound units) into phrases, sentences, and discourse raises a number of issues. Complexity can be investigated in each module independently but also between these modules, raising research questions such as whether the lesser complexity of a module will be balanced by the greater complexity of another.

Such a conception of EQUILIBRIUM can also be considered within a domain, if, for example, one attempts to check whether the greater complexity of the consonant or vowel system is counterbalanced by the lesser complexity of the syllabic structures (Maddieson 2011). Technically, comparing distinct domains such as phonology and morphosyntax is uneasy beyond simply counting elements, which, as remarked earlier, typically disregards the interactions between them.

On a more theoretical level, what is obviously missing from the relevant literature is the interactional complexity that arises from the division of labor and cooperation between different components, including members of the same module. However, it is not evident how many modules must ultimately be posited to account for how the production and interpretation of utterances work in a language. As a matter of fact, Lieberman (2012) goes as far as rejecting the idea of modules, arguing that the neurons of the brain are connected in a way similar to (though more complicated than) the parts of an automobile engine. However, if one subscribes to the modular architecture of language, it becomes important to understand how the modules interface with each other during the production and interpretation of utterances, certainly not in a linear way (McCawley 1998). According to the latter, the modules work concurrently rather than sequentially, as is made evident by, for instance, the correction of false starts while speaking and self-corrections of the interpretations of utterances as the discourse evolves.

Let's assume that the materials of a language fall in one or another module (viz., phonology, morphology, syntax, etc.), each of which makes a clear contribution to the overall system, while its components (such as individual sounds in a phonemic system) do not. We may then have to wonder whether languages do not fall in between complicated and complex systems, consistent with Loureiro-Porto and San Miguel's distinction (Chapter 8). The question is difficult to answer within the bit-complexity approach.

On the other hand, determining whether a language is complex or complicated becomes rather pointless without reference to something that it can be compared with. This explains why linguistic complexity has typically conjured up cross-linguistic research. Any measure of complexity presupposes or entails some cross-linguistic comparison of phonological inventories, morphological systems, and so on, as is evident, for instance, in discussions about whether or not creoles' grammars are simpler than those of other languages (e.g., McWhorter 2001, DeGraff 2001). This approach also suggests that a language may exist that has no, or very little, complexity built in it. However, from the point of view of the interaction of modules, McWhorter's (2001) and Gil's (2001, 2009) claims about the simplicity of the grammars of creole vernaculars and Riau, Indonesia, respectively, beg the question. However, see Gil's (2009) reaction discussed below.

1.3 From Static to Dynamic Linguistic Systems

The complexity of a linguistic system can be assessed synchronically, relative to a given time, regardless of what the system was like before. However, languages are constantly changing and being adapted to satisfy various communication pressures, including those that index speakers and the circumstances of their interactions. Beyond structures that may be assumed to be a static response to a fragile assemblage of structural constraints – in the spirit of a saying usually attributed to Ferdinand de Saussure, "la langue est un système où tout se tient" – linguistic systems are in a constant state of flux, with new components appearing and older ones evolving or disappearing. It thus makes sense to ask how complexity evolves under these ecological pressures, and see languages or their subsystems as *complex dynamical systems* (Bruckner et al. 2009).

Self-organization and *emergence* express how order and regularities arise from an initially chaotic state (as defined in Chaos Theory). They are fundamental processes in the study of physical and biological complex systems, for instance, how ants may build optimal paths between their nest and a food source, how microscopic dipoles can align to create magnetic domains, and how traders' activity at a market can result in macroscopic events such as economic bubbles or crashes. These concepts can be invoked to account for linguistic phenomena such as the emergence of new language varieties. For example, both SELF-ORGANIZATION and EMERGENCE can be invoked to explain how elements from several languages have been selected, in varying proportions, into a new variety, called creole. Linguistic systems can merge – with some features being selected and possibly modified, and others rejected – into what appears to be a new dynamic equilibrium.

Unlike in the science of complexity, linguistics stands out also by the limited attention that has been given to how complexity emerges, that is, from a diachronic perspective, relative to language development and to the phylogenetic emergence of language. Exceptions include Wang et al. (2004), Givón (2009), Lee et al. (2009), Mufwene (2012), and some of the authors contributing to this volume, especially Bart de Boer, P. Thomas Schoenemann, and Luc Steels, the first and the last based on modeling the emergence of language. It is to this fold of linguistic complexity that they were invited to contribute.

Different time scales relate to different evolutionary processes: from the ontogeny of complexity in child language acquisition, to its modifications in language change and the evolution of languages, to its rise during the phylogenetic emergence of Language. In ontogenetic and phylogenetic evolution, the focus is especially on how a system develops/evolves from architecturally poorer to richer structures (see, e.g., Dahl 2004, Givón 2009, Givón & Shibatani, eds. 2009). Regarding historical language changes, phenomena such as grammaticalization can be studied with a focus on whether they increase or decrease the complexity of the system (e.g., Heine & Kuteva 2007). Other attempts yet derive diachronic models from synchronic constraints, and observe how system coherence and complexity evolve between lower and upper bounds (Coupé et al. 2009). From a phylogenetic point of view, one should not dodge the question of how complexity arose during the transitions from vocalizations to naming and the rise of phonetic systems, to predication and the emergence of simple sentences, all the way to modern linguistic systems (e.g., Mufwene 2013).

The linguistics discourse has generally overlooked the dynamics of the linguistic elements in relation to each other, such as what may happen when a new sound is added to the phonetic inventory of a language; or when a preposition is used as the syntactic head of the predicate phrase (like in *dis buk fuh you* 'this book [is] for you' in Gullah), whereas a verb has traditionally been required in this position in English (Mufwene 1996). That is, while speakers/signers modify the extant system with their innovations, the latter may trigger other adjustments in the system. This is the case in the sentence *You bin fuh come* 'you had/were expected to come' in Gullah, where, because it can function as head of a predicate phrase, the preposition *fuh* has also been coopted as a marker of OBLIGATION, the counterpart of a modal verb in English. (As head of the predicate phrase it can also be modified by the anterior tense marker *bin*, regardless of whether it functions as a preposition or as a modal marker.) Such a change by cooption of extant materials is undoubtedly true of other cultural systems, which are also adaptive but depend primarily on the activities that shape them, those of the practitioners of the culture. Future research should return to this issue, which arises also from some of the contributions to the present book.

An important question in such systemic adaptations is: What are the forces or constraints responsible for linguistic change? Answering this question offers complementary and enriching perspectives on linguistic complexity. Indeed, the previous approaches can all fit a framework where linguistic structures are considered in isolation and studied on the basis of their internal (possibly dynamic) patterns of occurrences or interactions. But considering the various dimensions – social, cognitive, and pragmatic – of what may be called the *ecology of language* (Mufwene 2001, 2008; Coupé 2016) opens further avenues toward a more complete understanding of linguistic complexity.

1.4 Complexity and Language Ecology

Language ecology has usually been invoked in relation to social factors (Mufwene 2001, Lupyan & Dale 2010).[5] Indeed, a language does not exist outside its social environment; it is a communal creation, with structures shaped through speakers' communicative acts. It displays emergent patterns, which linguists have attempted to capture in the form of rules and constraints, from a synchronic perspective. However, there are ecological factors that arise from within the system itself that also influence the evolution of a language (e.g., frequency, transparency, regularity, and length of particular variants). They determine which variants will prevail and which ones will remain minority alternatives or will be given up.[6] Innovations and their replications (or copies) compete among themselves, subject to these and other ecological factors, social and otherwise (Mufwene 2001, 2008; Blythe & Croft 2009). This is especially noticeable in cases of language contact, when a new variety (such as a creole) emerges and retains only a subset of the variants in the prevailing language (called *lexifier*) and only some of the competing substrate features are selected into the emergent language variety.

Equally, if not more, interesting are cases where the competition[7] is not resolved. For instance, in (standard) English, the primary stress in the word *exquisite* may be placed on the first or second syllable; a relative clause may start with a null complementizer, with the complementizer *that*, or with a

[5] Mufwene actually applies the term *ecology* to a wider range of factors, both internal and external to particular languages, some direct and others indirect, that influence the evolution of a language, including its vitality. He applies the term to any factor that may be considered as (part of the) environment relative to a language (variety) or a linguistic feature being discussed. Relative to language evolution, some ecological factors are economic and historical. Relative to the phylogenetic emergence of Language, Mufwene (2013) singles out the human anatomy and the brain/mind as critical ecological factors.

[6] Linguists such as Weinreich et al. (1968), McMahon (1994), and Labov (2001) have invoked actuation (similar to but not exactly the same as actuator in physics) in reference to the particular combination of factors, which are indeed ecological, that produce particular changes at specific places and at specific points in time. This tradition of course underscores the need to approach language evolution from the point of view of complexity and emergence, as these notions are construed in the science of complexity, in dynamical terms.

[7] As explained in Mufwene (2008), "competition" is used here in the same sense as in *evolutionary biology*, applying to variants, organisms, or species that ecology may not sustain equally, favoring one or some but disadvantaging the other(s). In languages, variants for the same function (including languages spoken in the same community) are often rated differently by their speakers or signers, a state of affairs that explains why some disappear. From an evolutionary perspective, and even that of language ontogenetic development (influenced by who the learner interacts with), different speakers/signers may not rate the variants in identical ways.

relative pronoun (a *wh* form); and some speakers say *I want you not to come*, whereas others prefer *I want you to not come*. There are indeed a host of similar examples not only in English but in other languages too. What also appears evident in such cases is not only the number of distinctions that are made but also the ways in which the distributions of the competing expressions are articulated. No assessment of complexity in a language should ignore this interactive aspect of the system, which is consistent with Saussure's notion of OPPOSITION between forms or between constructions: the expressions derive their meanings from how they are opposed to or distinguished from each other.

How is variation managed in a language or speech community? Can it remain free, in the sense that a speaker/signer may use any variant or another without any communicative or social consequences? Or is it constrained by other factors that are social, such as age, gender, ethnicity, profession, and level of education, or those that stem from the precise context of interaction? Are the constraints rigid or flexible? The interfacing of systems (consisting of structural units and rules) and social constraints emanating from the communities using and shaping the languages appear to foster alternative interpretations of linguistic complexity, which also explains variation in the way that linguists discuss it, as is evident in, for instance, Sampson et al. (eds., 2009) and Massip-Bonet and Bastardas-Boada (eds., 2012). To the extent that languages can be construed as communal systems, complexity arises at least as much from the dynamics of interaction within the population associated with the language, as from the actual system hypothesized by the linguist (or any analyst). Linguistic complexity therefore conjures up complexity of linguistic structures and external constraints exercised by ecological factors, including specific kinds of social interactions and the particular business or social networks in which one operates.

No speaker/signer has complete knowledge of their communal language as an ensemble of idiolects (Mufwene 2001), while they all use it and adapt their respective idiolects relative to other users. The speakers'/signers' mutual accommodations and their respective responses to novel communicative pressures (which are similar to adaptive responses of elements of better understood complex adaptive systems) drive change or evolution. This peculiarity explains the claim that languages as both practices and systems are in a constant state of flux, hardly staying in equilibrium, and are therefore emergent phenomena.

These dynamic aspects of complexity are hardly quantifiable. They also make it obvious that, as stated by Beckner et al. (2009), the agents of the emergence of complexity are the speakers/signers who manipulate the system. They are the ones that modify it, innovate new forms and structures, introduce new dynamics of interaction among the different components of the system on different levels, and therefore modify the patterns of complexity in one way or another. On the other hand, this agency also sets the discourse on linguistic