A Report on the Workshop on Complexity in Language: Developmental and Evolutionary Perspectives

Tao Gong & Christophe Coupé

1. Overview

Complexity can be viewed as “the property of a real world system that is manifest in the inability of any one formalism being adequate to capture all its properties” (Mikulecky 2001: 344). In the past few decades, this notion has raised significant interest in many disciplines, from physics to biology, mathematics, artificial intelligence, etc. (Waldrop 1992, Simon 1996, Dahl 2004, Gell-Mann 2005, Hawkins 2005, Friederici et al. 2006, Risager 2006, Boogert et al. 2008, Larsen-Freeman 2008, Liu 2008, Riecker et al. 2008, Lee et al. 2009, Faracas & Klein 2009, Givón 2009, Mitchell 2009, Pellegrino et al. 2009, Cyran 2010, Trudgill 2011, and McWhorter 2011, among others); more recently, this cross-disciplinary endeavor has reached linguistics, and scholars of various theoretical backgrounds have been keen to test its relevance to language (e.g., Gibson 1988, Changizi & Shimoj 2005, Papagno & Cecchetto 2006, Lee et al. 2007, Suh et al. 2007, Miestamo et al. 2008, Givón & Shibatani 2009, Sampson et al. 2009). However, it is still unclear what complexity actually means in linguistics, what yardstick can be used to measure complexity, especially in comparing language varieties, and what conceptions are relevant to accounts of structures and functions of languages.

For the purposes of fostering a dialog between scholars of diverse but complementary backgrounds on these topics, Salikoko S. Mufwene, then a fellow at the Collegium de Lyon, in collaboration with researchers at the Laboratoire “Dynamique du Langage” at the Université de Lyon, convened a workshop on Complexity in Language: Developmental and Evolutionary Perspectives on 23–24 May 2011. Participants included linguists, anthropologists, statistical physicists (modeling communal aspects of language), computer scientists, and mathematicians. Most of them agree on seeing language as a complex adaptive system as described by Steels (2000) and Beckner et al. (2009). According to this view, a linguistic system involves a number of interacting units and modules that generate structural and interactional complexity on several levels. Meanwhile, these scholars, based on their expertise, have distinct research foci on linguistic complexity. The workshop consisted of 14 talks touching upon various topics related to linguistic complexity, such as (i) where linguistic complexity lies, (ii) how it emerges and evolves ontogenetically or phylogenetically, and (iii) how it is measured using...
approaches adopted and adapted from disciplines other than linguistics. In this review, we will go over these talks and present our opinions on future research of linguistic complexity.

2. Describing Complexity in Linguistics

Among the participants, linguists presented their complementary theories of linguistic complexity. The organizer of the workshop, Salikoko S. Mufwene (University of Chicago & Collegium de Lyon), divided linguistic complexity into (i) complexity within a communal language, which deals with the dynamics that produce communal norms, (ii) bit complexity, which reflects the number of units and rules in the lexicon, syntax, phonology, and other linguistic modules, and (iii) interactional complexity, which refers to the interactions of units and rules within their respective modules and of latter with one another. He pointed out that language evolved as a communal technology for mapping conceptual structures onto physical structures, and that all these forms of complexity emerged due to ecological and social factors (Mufwene, in press), interaction constraints and self-organization (Camazine et al. 2001).

William Croft (University of New Mexico) viewed linguistic complexity as structural complexity existing in modern languages and evolutionary complexity echoing the increasing structuration that led to modern languages. He claimed that the selective pressure for structural complexity came from the necessity of establishing common ground in joint actions (Bratman 1992, Tomasello et al. 2005), and that it was human social cognitive abilities that helped build up such common ground. Based on the evidence from language acquisition and the evolution of semasiographic systems such as images, notations and writing, he summarized the key features of the evolution of social cognition in humans, including gradualness, context-dependency and multimodality, which further inspired some speculations on evolutionary complexity in language.

William S-Y. Wang (Chinese University of Hong Kong) argued that language was a diffusive and heterogeneous system. Socio-cultural reasons drove the evolution of linguistic complexity, and therefore measuring current linguistic complexity, such as phonological complexity, could shed light on some age-old controversies regarding past linguistic communities, such as whether language origin was a monogenetic (Atkinson 2011) or polygenetic (Freedman & Wang 1996, Coupé & Hombert 2005). He also discussed one outcome of linguistic complexification, namely lexical and construction ambiguities. He argued that cross-language studies of these ambiguities could yield important insights into linguistic and cognitive universals.

Barbara Davis (University of Texas at Austin & Collegium de Lyon) proposed a biological-functional perspective to phonological acquisition, viewing the acquisition of phonological complexity as a consequence of interactions of biological and social components of language to achieve maximal functional efficiency (Davis et al. 2002). She introduced frame-content theory (MacNeiagle 1998), which follows this perspective and aims to explain the acquisition of one type of phonological complexity, namely the C(onsonant)V(owel) co-occurrence patterns in the world’s languages. This theory is supported by the experimental results of
English or Korean learning (Lee et al. 2010) infants, as well as the acquisition data of other languages.

Albert Bastardas-i-Boada (University of Barcelona) generalized a philosophical, holistic view of language contact. He conceived of an ecosystem of language, including brain/mind, social interaction, group, economics, media, and political factors. All these dynamic factors co-produced and co-determined the forms, usage, and evolution of language.

Unlike linguists, the anthropologist Thomas Schoenemann (Indiana University) focused on complexity in the physical substrate of language (the human brains) and of human behaviors. He advocated the theory of language-brain coevolution (Deacon 1997), and suggested that an increasing complexity of hominin conceptual understanding led to an increasing need for syntax and grammar to fulfill efficient communications (Schoenemann 1999). He argued that concepts were based upon networks connecting different brain regions, and that the size of those regions across species was proportional to the degree of elaboration of the functions they underlie. In the past, the increase in brain size was correlated with the increase in degree of specialization of parts of the brain (Schoenemann 2006).

The genetic linguist Jean-Marie Hombert (University of Lyon 2) focused on the relation between population size, social complexity and language dispersal. Based on genetic and demographic data, he suggested that Pygmy hunter-gatherers and Bantu-speaking farmers in Central Africa shared a common ancestry (Quintana-Murci et al. 2008, Berniell-Lee et al. 2009). This case study illustrated that population size and hierarchy could be two important factors within linguistic communities that helped develop linguistic complexity.

3. Measuring Complexity in Linguistics

Apart from describing and circumscribing linguistic complexity, many talks tried to propose general procedures or quantitative measures to evaluate different aspects of linguistic complexity. Artificial intelligence expert Luc Steels (University of Brussels & Sony CSL, Paris) presented a general procedure to account for linguistic complexity. This procedure includes five steps: (i) describing a complex linguistic structure, (ii) identifying its function, (iii) reconstructing processing and acquiring mechanisms for this structure, (iv) surveying its variations in languages, and (v) identifying its selective advantage. Such an approach helped pinpoint the different factors that contributed to linguistic complexity. In addition, Steels presented several simulation studies that explored the evolution of complexity in semantics and syntax. These studies supported his recruitment theory of language evolution (Steels 2009), stating that (i) strategies and structures that could satisfy communicative needs, reduce cognitive efforts, and increase social coherence could be adopted by language users and survive in languages, and (ii) the emergence of linguistic complexity was a process of self-organization of existing systems and of recruitment of new mechanisms in a cultural environment.

Statistical physicist Vittorio Loreto (Sapienza University of Rome & Institute for Scientific Interchange, Torino) pointed out that statistical physics served
as an efficient means to study linguistic dynamics and complexity (Loreto et al. 2011). Relying on language game simulations (Baronchelli et al. 2006, 2010, Puglisi et al. 2008), he argued that this approach could help understand: (i) How collective behaviors (e.g., common lexical items or linguistic categorization patterns) originated in local interactions, (ii) what were the minimum requirements for a shared linguistic feature to emerge and diffuse, (iii) how to examine asymptotic states in language evolution, and (iv) what roles population size and topology played in language evolution.

Mathematician Ramon Ferrer-i-Cancho (Universitat Politècnica de Catalunya) analyzed the effect of two quantifiable constraints on the word order bias in languages, namely, the predictability of a sequence of words and the amount of online memory for handling the head-modifier dependencies (Ferrer-i-Cancho 2008). The results obtained from this mathematical analysis, and also observed in simulation studies (e.g., Gong et al. 2009), illuminated empirical findings (e.g., Dryer 2008) and inspired further discussions (e.g., Cysouw 2008).

Linguist Lucía Loureiro-Porto (University of Palma de Majorca), collaborating with statistical physicist Maxi San Miguel and Xavier Castelló, defined two quantitative parameters, social prestige of different languages and individual volatility (speakers’ willingness to shift their current language to another), to examine the effect of social complexity on language competition. Using agent-based modeling and two sets of abstract equations of language competition (Abram & Strogatz 2003, Minett & Wang 2008), these scholars compared language competition in different social networks, and observed that (i) volatility was more powerful than prestige to cause language death and (ii) bilingualism accelerated language death (Castelló et al. 2008).

Apart from artificial simulations, evolutionary linguist Bart de Boer (University of Amsterdam) argued that the cultural learning paradigm (Galantucci 2005, Scott-Phillips & Kirby 2010) could help distinguish the effect of cultural learning from the effect of cognitive biases on linguistic complexity. His case experiment of whistle transmission in chains of human subjects revealed that the emergence of complex combinatorial structures was mainly due to cultural learning, with only limited influence from cognitive biases.

Psycholinguist Fermin Moscoso del Prado (University of Lyon 2) adopted the general framework of information theory and applied Gell-Mann’s (1995) notion of ‘effective complexity’ to language. Accordingly, the complexity of a linguistic system could be reflected by the length of the most compact grammar that describes the structural regularities of this system. He showed how to mathematically apply this approach to large text corpora, and how different linguistic components — lexicon, syntax, pragmatics or morphology — could be evaluated independently. A comparison between English and Tok Pisin corpus indicated that it was erroneous to claim that creole/pidgin grammars are simpler (McWhorter 2001).

Linguists Christophe Coupé, Egidio Marsico, and François Pellegrino (University of Lyon 2) concentrated on phonological systems and proposed a quantitative approach to analyze their complexity. Based on a genetic linguistics balanced dataset of 451 phonological inventories, namely the UPSID database (Maddieson 1984, Maddieson & Precoda 1990), they measured the strength of
interactions between phonemes, and suggested that the degree of complexity of the inventories was actually quite low. An evolutionary model was then derived from the synchronic data in an effort to further assess the extent to which the structure of the inventories could be understood and predicted (Coupé et al. 2009).

4. Future Research of Complexity in Linguistics

The workshop gathered many state-of-the-art studies on linguistic complexity, and offered opportunities for interested scholars to exchange ideas, methods, and findings across research areas and disciplines. It provided several important guidelines for the future research in linguistic complexity. First, complexity in linguistics is manifest in many aspects, including not only linguistic structures, but also population interactional dynamics, and cultural environments. As for the linguistic structures, variation and diversity of languages provide a rich repertoire of phenomena, which should be considered when we devise general theories of structural complexity (Evans & Levinson 2009). To this end, the typological database, namely the World Atlas of Language Structures (WALS, see http://wals.info), which records different types of structural variations across many of the world’s languages, serves as an important resource for future research.

As for the language users, neurolinguistic and psycholinguistic research, which examines empirical bases of linguistic behaviors in the human brain and traces individual differences in language acquisition and processing, will bear significantly on the embodied aspects of linguistic complexity. Meanwhile, structural complexity reflects conceptual/cognitive complexity. Examining structural complexity could help us better understand the Sapir–Whorf Hypothesis (Sapir 1929, Whorf 1940) and discuss how linguistic structures and usage influence human thoughts and non-linguistic behaviors.

As for the cultural environment, research on language contact from historical linguistics, sociolinguistics, population-based studies (e.g., Mufwene 2001, 2006, Ansaldo 2009) as well as simulations (e.g., Steels 2000, Brighton et al. 2005, Gong 2010) will yield useful insights on how interactions and cultural variations affect linguistic complexity, and vice versa.

Regarding these various approaches, challenge remains to cross the gap between quantitative approaches and wider and more conceptual notions, such as bit complexity, evolutionary complexity, or structural complexity — to recall only a few mentioned earlier. Quantitative approaches may provide figures, but these figures sometimes fail to necessarily uncover the true mechanisms at hand. Meanwhile, conceptual notions are instructive, but sometimes these notions suffer from a lack of empirical studies to support them. Therefore, revisiting earlier theories with the vocabulary and concepts of complexity theory is undoubtedly useful to better frame intricate phenomena, and further articulation with smaller scale aspects could be even more precious.

Second, a significant question concerning linguistic complexity that requires further investigation is the degree to which all languages are equally complex, or, whether there is compensation among linguistic components, say,
whether a language with a rich morphology tends to exhibit a simple phonology or syntax. Such assumptions are found in most introductory textbooks to linguistics, yet there are very few attempts to provide strong arguments for or against them. A reason for this is the difficulty in coming with complexity measures that can address the various linguistic components, such as lexicon, syntax, phonology, morphology, and so on in an integrative manner. Most current studies are usually confined to one of these domains, and need to be revised to reach beyond. To this end, databases like WALS may once again come in handy, and corpora or entropy based approaches or measures (such as Fermin Moscoso del Prado’s) seem promising for the future research.

Third, unlike previous theories that relied upon biological evolution to explain the origin of language and the evolution of linguistic universals (e.g., Pinker & Bloom 1990), modern theories pay great attention to cognitive abilities in humans and cultural processes in which language is acquired and transmitted. Language is inseparable from its socio-cultural environment and cultural evolution is too rapid for biological evolution to fix adaptations to arbitrary features of language (Christiansen & Chater 2008). Therefore, many universal properties of language should be ascribed to general cognition and cultural evolution (Evans & Levinson 2009, Dunn et al. 2011). The different angles adopted by many talks in the workshop to describe and explain linguistic complexity — from joint action, shared intentionality, brain-language co-evolution, individual processing and memory constraints, to human migration, population size and hierarchy, social networks and cultural learning — are actually falling into these two perspectives. Incorporating these perspectives into linguistics will greatly change the nature of this discipline (Levinson & Evans 2010).

Finally, it seems that no single discipline can alone account for all aspects of linguistic complexity. On the one hand, although linguists can carefully record in detail different types of variations in modern languages, powerful methods are needed to bring light to the correlations hidden in surface structures, to notice the selective pressures cast by other relevant factors, or to reconstruct the evolutionary trajectories leading to those variations.

On the other hand, although the research methods from other disciplines, such as genetics, anthropology, statistical physics, mathematics, and computer modeling, can quantitatively shed light on aspects of linguistic complexity, without sufficient guidance from linguistics, studies adopting those methods may pay unjustified attention to trivial factors or overlook more significant ones. For example, Atkinson’s (2011) mathematical analysis based on the phonemic diversity in languages was questioned by some of the workshop attendees for disregarding the influence of population size or language contact. Dunn et al.’s (2011) approach to word-order typology, inspired by evolutionary biology, was also criticized for ignoring the powerful effect of contact on typological change.

Therefore, questions concerning linguistic complexity have to be tackled based on a multi-disciplinary approach, a prerequisite to making sense of seemingly contrary positions, providing alternative perspectives, and ruling out solutions plausible only in the framework of a single discipline. This approach could offer the best prospect of arriving at an adequate and comprehensive understanding of linguistic complexity (Bickerton & Szathmáry 2009).
References


Givón, Thomas. 2009. The Genesis of Syntactic Complexity: Diachrony, Ontogeny,
McWhorter, John H. 2001. The world’s simplest grammars are creole grammars. Linguistic Typology 5, 125–166.
Biolinguistics ★ Forum ★


Steels, Luc. 2009. Cognition and social dynamics play a major role in the for-

Tao Gong
University of Hong Kong
Department of Linguistics
Pokfulam Road
Hong Kong Island
Hong Kong
tgong@hku.hk

Christophe Coupé
CNRS & Université Lyon 2
Laboratoire Dynamique du Langage
ISH, 14 avenue Berthelot
69007 Lyon
France
christophe.coupe@ish-lyon.cnrs.fr