Advances in Complex Systems
Vol. 15, Nos. 3 & 4 (2012) 1203002 (9 pages)

DOI: 10.1142/S0219525912030026

World Scientific
www.worldscientific.com

LANGUAGE DYNAMICS

ANDREA BARONCHELLI

Department of Physics,
College of Computer and Information Sciences,
Bouvé College of Health Sciences,
Northeastern University,
Boston MA02120, USA
a.baronchelli@neu.edu

VITTORIO LORETO

Sapienza University of Rome, Physics Department,
Piazzale Aldo Moro 2, 00185, Roma, Italy

and

Institute for Scientific Interchange ISI,
Via Alas 11/c, 10126 Torino, Italy
vittorio.loreto@roma1.infn.it

FRANCESCA TRIA

Institute for Scientific Interchange ISI,
Via Alassio 11/c, 10126 Torino, Italy
tria@isi.it

Received 8 May 2012
Published 24 May 2012

Thirty authors of different disciplines, ranging from cognitive science and linguistics to mathematics and physics, address the topic of language origin and evolution. Language dynamics is investigated through an interdisciplinary effort, involving field and synthetic experiments, modelling and comparison of the theoretical predictions with empirical data. The result consists in new insights that significantly contribute to the ongoing debate on the origin and the evolution of language. In this Topical Issue the state of the art of this novel and fertile approach is reported by major experts of the field.

Keywords: "In silico" linguistics; cultural evolution; language games; experiments.

1. Introduction

This Topical Issue is devoted to the theme of Language Dynamics, a rapidly growing area that focuses on all the processes related to the emergence, evolution, change and extinction of languages. The idea of collecting in a single issue contributions coming from several disciplines aims to highlight the richness and variety of the different approaches to Language Dynamics by bringing together pieces of work

otherwise scattered in a very heterogeneous literature. The choice of Advances in Complex Systems shows on its turn the relevance that ideas and methods borrowed from statistical physics and complex systems science are progressively acquiring for studies on the origin and evolution of language. New theoretical and computational tools as well as synthetic modelling approaches have now reached sufficient maturity to contribute significantly to the current debate in Cognitive Science. In addition, the unprecedented advances in information and communication technologies are triggering a significant boost in the ongoing transition of linguistics into an experimental and computational discipline, where multiple evolutionary paths, timescales and dependences on the initial conditions can be effectively controlled and modeled. The combination of these two elements is opening terrific new opportunities for studying the emergence and evolution of languages, new communication and semiotic systems.

This journal issue is composed of fourteen contributions that mirror the effort of thirty researchers actively involved in the field of language dynamics. Their disciplines of belonging range from Linguistics, Cognitive Science, Psychology, Philosophy, Physics, Mathematics, Artificial Intelligence, Computer Science.

The contributions can be organized in three main sections devoted to: *data analysis*, *modelling* and *experimental approaches*. Of course, this classification is dictated solely by practical reasons, and it is (largely, perhaps) imprecise. Important overlaps among the three sections do in fact exist, and most of the research in language dynamics tends to blend several approaches together in order to gain a more solid and robust understanding. For instance, more and more often the modelling activity is paralleled by the comparison with empirical data. This helps to check whether the trends seen in real data are already compatible with plausible microscopic modelling of the individuals, or require additional insights. When coupled with a serious data analysis devoted to the discovery of emergent features, theoretical modelling can trigger a virtuous loop: the empirical measures inspire modelling schemes, the model analysis suggests new measures and observations, and the latter foster in their turn the evaluation and refinement of models.

Traditionally, the investigation of language dynamics suffered from the lack of extensive datasets mainly due to the difficulty of monitoring in a systematic and reproducible way the emergent steps of new linguistic features or new languages altogether. The situation has radically changed in the last few years. Unprecedented advances in information and communication technologies are enabling, for the first time, the possibility of precisely mapping the interactions, whether embodied and/or symbolic, of large numbers of actors, as well as the dynamics and transmission of information along social ties. From this point of view, the Web is of great help both as a platform to perform controlled online social experiments, and as a repository of empirical data on large-scale phenomena. In this way a virtuous cycle involving data collection, data analysis, modelling, and predictions can be activated, giving rise to an ever more rigorous and focused approach to language dynamics.

In the next sections we briefly describe the fourteen contributions comprising this issue.

2. Data Analysis

Three papers present the outcome of thorough analyses of existing datasets, focusing in particular on World Color Survey [3], the Automated Similarity Judgment Project data base [5], the World Atlas of Language Structures [4] as well as two datasets of word associations, namely the Human Brain Cloud database and the University of South Florida Free Association Norms database [10].

In particular, in his paper *Power Laws and Other Heavy-tailed Distributions in Linguistic Typology*, Gerhard Jäger investigates the quantitative distribution of language types across languages of the world. The studies are based on three large-scale typological data bases: The World Color Survey, the Automated Similarity Judgment Project data base, and the World Atlas of Language Structures. The main finding is that a surprisingly large and varied collection of linguistic typologies display a self-similar behaviour whose signature is given by the existence of power-law tails for specific observables.

Again concerned with the World Colour Survey is the paper *Why are Basic Color Names "Basic"?* by Animesh Mukherjee, Vittorio Loreto and Francesca Tria. This article addresses the question of a quantitative definition of Basic Color Names. Though it is widely known that color names across the world's languages tend to be organized into a neat hierarchy with a small set of "basic names" featuring in a comparatively fixed order across linguistic societies, still no statistical definition that quantitatively separates the basic names from the rest of the color words across languages exists. Mukherjee *et al.* present a rigorous statistical analysis of the World Color Survey database, showing that those names for which a population of individuals show a larger overall agreement across languages turn out to be the basic ones. This finding is also in agreement with the color name hierarchy. It is interesting to notice that the origin of this hierarchy [6] and the universality of color naming patterns in the World Color Survey have been recently explained [2] in the modelling scheme of Category Game [12].

Finally the paper *Complex Structures and Semantics in Free Word Association* by Pietro Gravino, Vito D. P. Servedio, Alain Barrat and Vittorio Loreto, explores the outcomes of word associations experiments. The idea here is that of looking at word-associations structures as proxies, or reflects, of the underlying of our conceptual spaces. Gravino *et al.* investigate the directed and weighted complex network of two free word associations experiments: on the one hand the massive multiplayer web-based Word Association Game known as Human Brain Cloud, and on the other hand the South Florida Free Association Norms experiment. A particularly interesting result concerns the existence of a typical scale for the word association process, arguably related to specific conceptual contexts for each word. In addition, the authors validate their results against the WordNet semantics network [8, 9],

pointing out the basic cognitive mechanisms underlying word associations when they are represented as paths in an underlying semantic network.

3. Modelling

The papers in this section address a wide range of language-related questions by introducing suitable modelling schemes. The spectrum of the addressed issues ranges from the emergence of consensus in the appearance of linguistic features, resulting either from cultural or evolutionary dynamics, to language learning and language competition, and to the interaction between language and mind.

In their contribution *Multilevel Alignment Maintains Language Systematicity*, Remi van Trijp and Luc Steels investigate how a shared grammar can arise in a multi-agent population playing language games [15]. Their approach taps on a growing body of literature that looks at language as an evolving and self-organizing system, whose components are constantly (re)shaped by language users in order to maximize communicative success and expressive power, while at the same time minimizing the effort [14]. For instance the emergence of a common lexicon can be explained in the very simple framework of the Naming Game [1, 13]. Agents score all associations between words and meanings in their lexicons and update these preference scores based on communicative success. Thus, a positive feed-back loop between success and use arises, causing the spontaneous self-organization of a shared lexicon. In this Issue van Trijp and Steels describe how a population can arrive at a shared grammar, in which the same problem of variation arises because each agent invents and acquires their own grammatical constructions. In particular, they introduce the notion of multi-level alignment, consisting in an update of the agent's inventories of preference scores that is not restricted only to the constructions that were used in the utterance, but extends also downward and upward in the subsumption hierarchy.

In his contribution *Neutral Evolution: A Null Model for Language Dynamics*, Richard Blythe reviews the task of aligning simple models for language dynamics with relevant empirical data. Blythe proposes, in particular, the adoption of carefully constructed null models. He argues that rejection of a null model must have important consequences for theories about language dynamics if modelling is truly to be worthwhile. The main claim here is that the stochastic process of neutral evolution (also known as genetic drift or random copying) is a viable null model for language dynamics. To support this hypothesis Blythe surveys empirical evidence in favour and against neutral evolution as a mechanism behind historical language changes, highlighting the theoretical implications in each case.

The following two contributions are closely related. Both the paper *Language, Categorization, and Convention* by Louis Narens, Kimberly A. Jameson, Natalia L. Komarova and Sean Tauber and *Evolutionary Game Theoretical Model of the Evolution of the Concept of Hue, A Hue Structure, and Color Categorization in Novice and Stable Learners* by Ragnar Steingrimsson, address language dynamics

from an evolutionary game theory point of view [11]. Narens and colleagues investigate how linguistic meaning, threaded as a convention, can arise for color categories in populations of simulated agents. Their method resort to concepts from evolutionary game theory to devise a language game where agents assign names to color patches, under minimal assumptions about agents' perceptions and learning processes. Through various simulations they show that the evolutionary dynamics push populations to stationary equilibria, which can be interpreted as achieving shared population meaning systems. Steingrimsson's paper adopts evolutionary game theory to investigate how a color categorization can evolve, i.e., a finite partition of a hue circle in which perceptually similar hues are each represented by an icon chip and the circle by a finite number of icon chips. It is argued that, in order to produce a stable color categorization system, learners need neither to share the same icon chip structures, nor do these structures have to be fully developed for a population of learners.

The next contribution *Word and Category Learning in a Continuous Semantic Domain: Comparing Cross-Situational and Interactive Learning* by Tony Belpaeme addresses the problem of how young learners acquire the meaning of words, a fundamental point to language development and cognition. Belpaeme presents two models that acquire meaning-word mappings in a continuous semantic space. The first model is a cross-situational learning model in which the learner induces word-meaning mapping in an unsupervised manner through repeated exposures. The second model is a social model, in which the learner and teacher engage in a dyadic learning interaction to transfer word-meaning mappings. It is shown that while cross-situational learning turns out to be viable, it is outperformed by social learning both in speed of acquisition and performance. Thus, while cross-situational learning is efficient for situations where referential ambiguity is limited, in more complex situations social learning is the more optimal strategy.

The contribution *Modelling Two-Language Competition Dynamics* by M. Patriarca, X. Castello, J. R. Uriarte, V. M. Eguiluz and M. San Miguel addresses the slightly different problem of language competition. The difference is due to the fact that such an issue unavoidably imposes the necessity of dealing with multiple populations. Indeed, models of language evolution usually focus on a single population, which is described as if it was isolated from the rest of the world. However, real populations are obviously not isolated, and keep interacting with each other. These steady contact between people play a major role in the evolution of languages. During the last decade, researchers in complex systems have devoted attention to the important area of language competition, studying how the fraction of speakers of several competing languages evolve in time. In this paper Patriarca and colleagues review recent advances in this direction and focus on three aspects. They first consider the shift from two-state models to three state models that include the possibility of bilingual individuals. The understanding of the role played by bilingualism is essential in sociolinguistics, in particular to clarify whether bilingualism

facilitates the coexistence of languages. Then, they analyze the effect of social interaction networks and physical barriers. Finally, Patriarca and colleagues show how to analyze the issue of bilingualism from a game theoretical perspective.

The final contribution of this section, *Studying the Impact of Language on the Mind by Constructing Robots that Have Language* by Domenico Parisi, discusses how language influences the mind by comparing robots that have language with robots that do not have language. Parisi observes that robots with language respond more adaptively to objects belonging to different categories and requiring different behaviours, and he notes that it is possible to show that categories of objects are represented differently in the neural network which controls the behaviour of the two types of robots. By exposing the robots to sounds which co-vary systematically with specific aspects of their experience, the robots can distinguish nouns from verbs and can respond appropriately to simple noun-verb sentences. Finally Parisi discusses how robotic experiments could be used to compare animals and human beings exposed to the task of developing a mental (in the robotic case neural) model of their environment which incorporates the co-variations among different aspects of their experiences.

4. Experiments

The four papers in this final section represent examples of a rapidly growing trend in language dynamics, that of conceiving and performing reproducible experiments involving human beings to elucidate open questions in language dynamics and provide the community at large with corpora of empirical data. This trend, also dubbed Experimental Semiotics (ES), is being fuelled by the above-mentioned recent advances in the information and communication technologies as well as by the new analytic tools and modelling methods developed in the last few years. As was the case with biology, the combination of this two elements is triggering a significant boost in the ongoing transition of linguistics into an experimental discipline, where multiple evolutionary paths, timescales and dependences on the initial conditions can be effectively controlled and modelled. From this perspective, it is very likely that the new social platforms appearing on the Web (see for instance http://www.xtribe.eu), could rapidly become a very interesting laboratory for social sciences in general and for studies on language emergence and evolution in particular.

In their paper *Experimental Semiotics: An Engine of Discovery for Understanding Human Communication*, Bruno Galantucci and Gareth Roberts review how the recent growth of Experimental Semiotics offers a new option to investigate human communication. In order to introduce Experimental Semiotics, Galantucci and Roberts present results from three themes of research which emerged within it. In particular they illustrate how ES can provide an engine of discovery for understanding human communication. In fact, in complementing the other options, ES offers us unique opportunities to test assumptions about communicative behaviour,

both through the experimenters' planned manipulations and through the unexpected behaviours humans exhibit in experimental settings.

Along the same line, the paper *Language Dynamics in Structured Form and Meaning Spaces*, by Bart de Boer and Tessa Verhoef, explores how the structure of form and meaning spaces influences the nature and the dynamics of the form-meaning mappings in language. They do so by illustrating an experiment in which participants have to learn a set of complex form-meaning mappings that have been produced by a previous generation of participants. In general, in a structured form or meaning space not all forms and meanings are equivalent: some forms and some meanings are more easily confused with each other than with other forms or meanings. The paper argues that some fundamental properties of human language, such as discreteness and combinatorial structure as well as universals of sound systems of human languages follow from optimal communication in structured form and meaning spaces. In addition de Boer and Verhoef show how the experimental setting allows to investigate the relative contribution of functional constraints and of cognitive constraints.

Another example of game/experiment is presented in the paper *Experimental Approach to Speakers Perspectives and Dynamic Patterns in a Paired Wall Game* by Ryoko Uno, Keisuke Suzuki and Takashi Ikegami. Here the authors introduce a new interactive *wall game* to study the emergence of rules and symbols in communication dynamics. In this game, two human players alternately configure a pattern on a board to communicate with each other, with no explicit game scores or task to optimize. The main outcome of this experiment is the observation that subjects mainly communicated in two modes: a dynamic mode where players proceed through the game without assigning any meanings to the pattern, and a metaphoric mode, where players process with narrative reflection. Since one of the main features of grammar is to display subjects' intentionality in a systematic way, it is argued that the switching between the two modes observed in their experiment can be taken as a grammatical aspect that emerged in the process. The switching behaviour should be seen as a process that expresses a player's internal state using the medium (in this case, the patterns in the wall game), and a player's exploration of the medium is a necessary step before generating a grammar structure.

The final paper of this section *Understanding the Origins of Morphological Diversity: The Linguistic Niche Hypothesis*, by Rick Dale and Gary Lupyan, explores the interplay between expressive capacity and diversity in human languages. Recent evidence suggests that language structures are shaped by the social environment in which they are learned and used. In particular Lupyan and Dale themselves [7] demonstrated in an analysis of over 2000 languages that social structure (e.g., population size) significantly predicts the complexity of inflectional morphology. Languages spoken by smaller populations tend to employ more developed inflectional systems (e.g., verb conjugation). Languages spoken by larger populations tend to disprefer complex morphological paradigms, using lexical constructions instead. This relationship may exist because of how language learning

takes place in these different social contexts. In a smaller population, a tightly-knit social group in which infants are learning the local language permits accumulation of complex inflectional forms. In larger populations, adult language learning and more extensive cross-group interaction creates different pressures. In the current paper, Lupyan and Dale explore this learning-based hypothesis in two ways. First, developing an agent-based simulation that serves as an existence proof: as adult interaction increases, languages lose inflections. Second, they carry out a human study showing that adults who have more interaction with non-native speakers tend to take on different metalinguistic strategies. The results of the simulation and experiment lend support to the linguistic niche hypothesis: learning constraints, mediated through social structure, help to explain linguistic diversity.

5. Conclusions

We hope this Topical Issue will contribute to the ongoing process of creating an interdisciplinary community interested in Language Dynamics. We strongly believe that a cohesive community developing and sharing new tools and experimental protocols is a necessary condition for fast-paced progress in the fascinating area related to the origins and evolution of language.

Acknowledgments

We are grateful to Frank Schweitzer for stimulating us to realize this Topical Issue as well as for his continuous support. We also wish to thank Alvin Chong for his help and patience during the whole editorial process. Finally we warmly thank all the authors who kindly accepted to contribute this journal issue with their original point of view on Language Dynamics.

References

[1] Baronchelli, A., Felici, M., Caglioti, E., Loreto, V. and Steels, L., Sharp transition towards shared vocabularies in multi-agent systems, *Journal of Statistical Mechanics* **P06014** (2006).

[2] Baronchelli, A., Gong, T., Puglisi, A. and Loreto, V., Modelling the emergence of universality in color naming patterns, *PNAS* **107** (2010) 2403–2407.

[3] Cook, R., Kay, P. and Regier, T., The World Color Survey database: history and use, *Handbook of Categorisation in the Cognitive Sciences. Amsterdam and London: Elsevier* (2005).

[4] Haspelmath, M., *The World Atlas Of Language Structures*, Oxford linguistics (Oxford University Press, 2005).

[5] Holmann, E. W., Wichmann, S., Brown, C. H., Velupillai, V., Muller, A. and Bakker, D., Explorations in automated language classification, *Folia Linguistica* **42** (2008) 331–354.

[6] Loreto, V., Mukherjee, A. and Tria, F., On the origin of the hierarchy of color names, *PNAS* **109** (2012) 6819–6824.

[7] Lupyan, G. and Dale, R., Language structure is partly determined by social structure, *PLoS ONE* **5** (2010) e8559.

[8] Miller, G. A. and Fellbaum, C., *WordNet: An electronic lexical database* (MIT Press, Cambridge, MA, 1998).
[9] Miller, G. A., Wordnet — about us, http://wordnet.princeton.edu.
[10] Nelson, D., McEvoy, C. and Schreiber, T., The university of south florida free association, rhyme, and word fragment norms, *Behavior Research Methods* **36** (2004) 402–407.
[11] Nowak, M., *Evolutionary Dynamics: Exploring the Equations of Life* (Belknap Press, 2006).
[12] Puglisi, A., Baronchelli, A. and Loreto, V., Cultural route to the emergence of linguistic categories, *PNAS* **105** (2008) 7936.
[13] Steels, L., A self-organizing spatial vocabulary, *Artificial Life* **2** (1995) 319–332.
[14] Steels, L., Language as a complex adaptive system, in *Parallel Problem Solving from Nature PPSN VI* (Springer, 2000), pp. 17–26.
[15] Wittgenstein, L., *Philosophical Investigations* (*Translated by Anscombe, G.E.M.*) (Basil Blackwell, Oxford, UK, 1953).