

Emergence and Evolution of Natural Languages: New Mathematical & Algorithmic Perspectives

Edward G. Belaga

February 15, 2011

In their recent influential paper *Language Evolution : The Hardest Problem In Science ?* [1], Morten H. Christiansen et Simon Kirby wrote: “*Language is one of the hallmarks of the human species – an important part of what makes us human. Yet, despite a staggering growth in our scientific knowledge about the origin of life, the universe and (almost) everything else that we have seen fit to ponder, we know comparatively little about how our unique ability for language originated and evolved into the complex linguistic systems we use today. Why might this be ?*”

In the search of new approaches to *the problem of emergence and evolution of natural languages*, Mathematics, Theoretical Computer Science, as well as Molecular Biology and Neuroscience, both deeply penetrated and profoundly inspired by concepts originated in Mathematics and Computer Science, represent today the richest pools of formal concepts, structures, and methods to borrow and to adapt.

The mathematical and computational methods, both well-known [2] and new ones [3], have both the potential and good perspectives to be used in the automatic treatment of natural languages [4], viewed as formal structures, and in the formal simulation of their emergence and evolution, free from linguistic intricacies of the historical perspective.

On the other hand, it has been shown that such a perspective could be taken into account with the help of mathematical and computational methods common for modern biological evolutionary studies. In particular, these methods have permitted recently to shed a new light [5] into the origins of the Indo-European language family, “*the most intensively studied, yet still most recalcitrant, problem of historical linguistics*” [6]. As to the Neuroscience end of this vast research enterprise, one looks here for “*the action origins of language, via the Mirror Neuron System*” [7].

The approach which we are advancing in the present study is building on some implications of the well-known but almost universally disregarded semantically meaningful tight combinatorial and topological morphological structure of the verbal system of Semitic languages, and in particular, of Biblical Hebrew [8], to show that the corresponding linguistic fossils testify to the existence of a now extinct Proto-Language whose extremely tight verbal organization and mean-

ingful architecture made it both structurally strikingly similar and expressively vastly superior to humanly designed Assembler languages [9], – an absolutely novel, paradoxical phenomenon, never before and nowhere else observed and crying out for new explanatory linguistic paradigms.

Besides its intrinsic value, our approach has the obvious potential to both validate “*action to language*” Neuroscience conjectures, by suggesting their novel and authentic interpretation, and to provide a new and, for that matter, extremely rich testing ground for the aforementioned mathematical and computational methods of simulation and analysis of natural languages.

References

- [1] M. Christiansen, S. Kirby, Language Evolution : The Hardest Problem In Science ? In: Language Evolution, eds. M. Christiansen, S. Kirby, Oxford University Press, pp. 1-15 (2003).
- [2] A. Biermann, B. Ballard [1980] : Toward Natural Language Computation. American Journal of Computational Linguistics **6**, 71-86 (1980).
- [3] L. Zadeh, Precisiated Natural Language –Toward a Radical Enlargement of the Role of Natural Languages in Information Processing, Decision and Control. In: Proceedings of the 9th International Conference on Neural Information Processing (ICONIP’0Z) , Vol. 1, eds. Lipo Wang, Jagath C. Rajapakse, Kunihiko Fukushima, Soo-Young Lee, and Xin Yao (2002).
- [4] A. Mehler, R. Köhler, Introduction: Machine Learning in a Semiotic Perspective. In: Aspects of Automatic Text Analysis, eds. A. Mehler, R. Köhler, Series: Studies in Fuzziness and Soft Computing **209**, Springer (2007).
- [5] R. Gray, Q. Atkinson, Language-tree divergence times support the Anatolian theory of Indo-European origin. Nature **426**, 435-439 (2003).
- [6] J. Diamond, P. Bellwood, Farmers and their languages: the first expansions. Science **300**, 597-603 (2003).
- [7] M. Arbib, ed., Action to Language via the Mirror Neuron System. Cambridge University Press (2006).
- [8] L. McFall, The Enigma of the Hebrew Verbal System: Solutions from Ewald to the Present Day, The Almond Press, Cambridge (1982).
- [9] M. Johnson. Assembly Language: For Real Programmers Only! Prentice Hall Computer Pub, Upper Saddle River, NJ (1993).

Edward G. BELAGA
Institut de Recherche Mathématique Avancée de Strasbourg
7 rue René Descartes, F-67084 Strasbourg Cedex, FRANCE
e-mail: *edward.belaga@math.u-strasbg.fr*