Chapter 70
Towards a Bio-Inspired Theoretical Linguistics to Model Man-Machine Communication

Gemma Bel-Enguix
CNRS-Aix-Marseille University, France

M. Dolores Jiménez-López
Universitat Rovira i Virgili, Spain

ABSTRACT

The article provides an overview of what could be a new biological-inspired linguistics. The authors discuss some reasons for attempting a more natural description of natural language, lying on new theories of molecular biology and their formalization within the area of theoretical computer science. The authors especially explore three bio-inspired models of computation –DNA computing, membrane computing and networks of evolutionary processors (NEPs) – and their possibilities for achieving a simpler, more natural, and mathematically consistent theoretical linguistics.

INTRODUCTION

We live surrounded by machines. Computers/robots are able to perform many human activities that facilitate our everyday life. As computers/robots become more omnipresent and capable of performing complex tasks, the importance of enabling non-expert users to interact with them has increased. Since people already have extensive communication skills through their own natural language, researchers believe that natural language can provide the most useful and efficient way for people to interact with computers. In fact, natural-language interaction with robots has emerged as a significant research area (Brooks et al., 2012; Deits et al., 2012; Krishnamurthy & Kollar, 2012; Matuszet et al., 2012; Oates et al., 2000; Scheutz et al., 2013, 2011; Scheutz & Eberhard, 2008; Tellex et al., 2011; Yang et al., 2013). The idea is clear: if computers can communicate only through artificial languages specifically designed for them, its use will be restricted to a small amount of people. On the contrary, if users are allowed to interact with machines by using natural language, robots will be easily accessible by everybody.

DOI: 10.4018/978-1-4666-6042-7.ch070

The precise extent to which is feasible the goal of a complete simulation of human language use is still an open question. Even though many researchers agree that a complete simulation is very difficult (maybe impossible) to be reached, it seems clear that knowledge of how humans use language can help in the design of efficient human-computer interactions. According to Zue (1997), "for information to be truly accessible to all anytime, anywhere, one must seriously address the problem of user interfaces. A promising solution to this problem is to impart human-like capabilities onto machines, so that they can speak and hear, just like the users with whom they need to interact." Maybe computers will never use language as humans do, but the analysis, formalization and implementation of human natural language processing can contribute to improve human computer interfaces.

The main problem in order to obtained machines that are able to speak is that we still without having a good theory about human use of natural language. Language is one of the most challenging issues that remain to be explained from the physiological and psychological point of view. As a complex system, their formal modeling and simulation present important difficulties. Many computational models try to explain natural language and many cognitive models try to explain how humans process language. However, models proposed up to now have not been able to give neither a coherent explanation of natural language nor a satisfactory computational model for its processing. If the scope is natural language, we need to cross traditional academic boundaries in order to solve the different problems related to language. There is a need of connecting and integrating several academic disciplines and technologies in the pursuit of a common task: the explanation of the human processing of natural language.

Disciplines that must be involved in the description, explanation and processing of language are, among others, *linguistics*, *theoretical computer science* and *biology*. Linguistics is described as the scientific study of natural languages, so, it is not possible to afford the topic of language without taking into consideration the theories that linguists have proposed up to now. Theoretical computer science, especially formal language theory, provides the tools needed to formalize the different theories about language. And, finally, biology –as a pilot science that have influenced many disciplines during the 20th century– can offer natural tools for describing languages that may improve theories and models in Natural Language Processing (NLP) –the subfield of Artificial Intelligence that concerns the automated generation and understanding of natural language.

The need for interdisciplinarity to afford natural language processing makes this article to be placed in the confluence of, at least, three research areas. In fact, we try to highlight the relevance of the *biological* approach via *formal/computational* methods for explaining *natural language processing*. In the models reviewed in this article, the main research interest is *natural language processing*, the main model comes from *biology* and *computer science* is the way to transfer biological concepts to formal linguistics. Therefore, our goal here is to briefly present different frameworks for approaching natural language with methods imported from biology and formalized by computer science.

BIOLOGICAL MODELS IN NATURAL LANGUAGE PROCESSING

As we have said, linguistics and natural language processing have still the challenge to understand how natural language is acquired, produced and processed. Up to now, linguistics has not been able to solve these challenges, partly, because of the fail in the models adopted. Indeed, it has been proofed that natural language does not fit in the classical Chomskyan hierarchy (Chomsky, 1956) of languages that are the base of the mathematical models for linguistics. Rewriting methods used in a