Biocomputing: an insight from linguistics

Gemma Bel-Enguix · M. Dolores Jiménez-López

Abstract This paper is placed in a formal framework in which the interdisciplinary study of natural language is conducted by integrating linguistics, computer science and biology. It provides an overview of the field of research, conveying the main biological ideas that have influenced research in linguistics. Our work highlights the main methods of molecular computing that have been applied to the processing and study of the structure of natural language: DNA computing, membrane computing and networks of evolutionary processors. Moreover, some new challenges and lines of research for the future are pointed out, that can provide important improvements in the understanding of natural language as a structure and a human capacity.

Keywords Molecular computing · Language evolution · Interdisciplinary research Linguistics-biology-computer science · Theoretical computer science · Formal languages

1 The triangle computer science-linguistics-biology

In the science of twenty-first century, as well as in the society and culture, two chief phenomena can be easily detected: *specialization* and *convergence*. On the one hand, the more human beings know, the more we have to restrict the field of our activities. On the other hand, the more we learn, the more we comprehend the similarity among underlying mechanisms in many states of nature. Consequently, although disciplines are more and more specialized, research areas become more and more interdisciplinary; traditional boundaries fall down in order to achieve a new understanding of both human capabilities and the behaviour of different types of life. Moreover, we know the reality can be approached from multiple perspectives—most of them complementary—that provide new views of the same phenomenon.

Within this interdisciplinary framework, this paper presents an example of research that has been developed in the shared space between the boundaries of three areas that may have crucial impact in the future of the society: *linguistics*, *biology* and *computer science* (see Fig. 1). The final goal of our research is *linguistics*; the main model comes from *biology*; and *computer science* is the way to transfer biological concepts to formal linguistics and, at the same time, it provides a new methodology of research.

Molecular biology has been the best and more increasingly developed area in the last 50 years, and it seems that most of the future goals of the humanity have something to do with the deciphering of DNA and genetic engineering. Currently, biology—especially molecular biology—has become a pilot science, so that many disciplines have formulated their theories under models taken from biology. Computer science has become almost a bio-inspired field thanks to the great development of some areas of natural computing, i.e. DNA computing. For linguistics, the relationship is even earlier, provided that the publication of Darwin's work *The origin of the species*, in 1859, was the first step towards an integrative view of language as a biological construct.

Computer science is not only transforming classical methods of science, but also our everyday life, by means of theoretical approaches and applications. Nowadays, in a process which is similar to the one undergone in biology,

G. Bel-Enguix (✉) · M. D. Jiménez-López
GRLMC-Research Group on Mathematical Linguistics,
Universitat Rovira i Virgili, Av. Catalunya 35,
43002 Tarragona, Spain
e-mail: gemma.belenguix@gmail.com

M. D. Jiménez-López
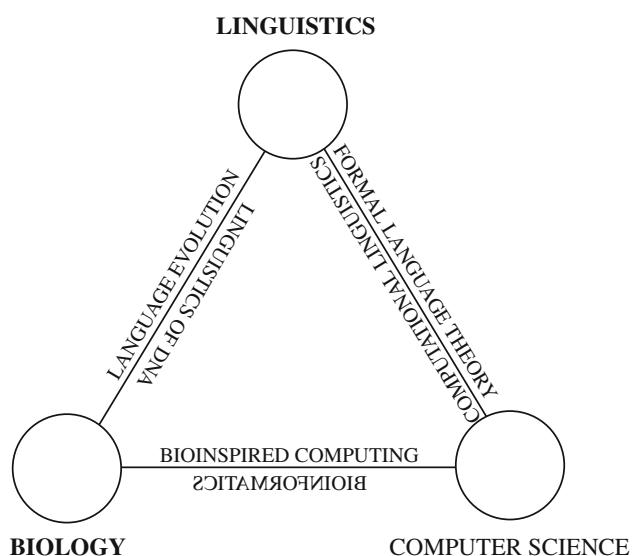e-mail: mariadolores.jimenez@urv.cat

Fig. 1 Interactions linguistics, biology, computer science

informatics provides to science revolutionary methods of research, models that can simulate unknown states of the nature and new perspectives in both, experimental and basic research.

Regarding linguistics, it still has the challenge to understand how natural language is acquired, produced and processed, in a multi-lingual society that needs some tools for linguistic interaction, automatic translation and human-computer interfaces. 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 proved that natural language does not fit in the classical Chomskyan hierarchy of languages, that constitutes the base of mathematical models for linguistics. Moreover, the metaphor of the mind as a computer seems to be almost exhausted. Finally, psychological models have a lack of formalization that prevents them to be conveniently implemented.

We think the time has arrived for starting a new paradigm in natural language research, when not only computer science, but also other disciplines, specially biology, can help linguistics by providing models, methods and fruitful interactions.

Surprisingly, and despite the fact that the genetic code has been considered a code from its discovery (Watson and Crick 1953), linguistics has not attempted to construct a new paradigm taking advantage of the great developments in molecular biology. Nevertheless, it is an especially suggestive idea, because of the interesting similarities between natural language and the genetic code in different levels.

Natural Language Processing (NLP) can take great advantage of the structural and "semantic" similarities between these codes. Specifically, taking the systemic code units and methods of combination of the genetic code, the methods of such entity can be translated to the study of natural language. Therefore, NLP could become another "bio-inspired" science, by means of theoretical computer science, that provides the theoretical tools and formalizations which are necessary for approaching such exchange of methodology.

Nowadays, having a look at the current situation of inter-disciplinarity inside the triangle, the two more successful connections are the pairs that integrate computer science, namely: *bioinformatics* and *computational linguistics*. In fact, computer science has lead molecular biology to the predominant situation it has today. Moreover, computational linguistics implies the use of computers in the analysis and study of large corpora and data bases of linguistic data. This has eased the arising of applied linguistics and language industries. The possibility of dealing with large amount of data is a great advance for science, and has solved one of the most important problems that both biology and linguistics had to cope with.

This article is chiefly concerned with theoretical approaches and methodological interaction: how biology is a model for computer science and linguistics, how some molecular structures can be understood in terms of codes, and how can these three disciplines collaborate to build common models and methodologies.

In this way, a theoretical framework can be drawn, where biology, linguistics and computer science exchange methods and interact. For linguistics, the bio-inspired computational paradigm can be a powerful tool for explaining language capacity from two different perspectives: *synchrony*—or the view of language as a complex system—and *diachrony*—or language understood as an evolutionary system.

Following the main ideas that have been explained here, we develop hereafter the state of the art and future lines of research within the interaction between the three disciplines: *computer science, biology, linguistics*. However, taking into account that bioinformatics (interaction computation-biology) and computational linguistics (interaction computation-linguistics) have been highly developed, we will focus on the interaction biology-linguistics both directly or through the methodological help of computer science.

The paper is structured as follows. In Sect. 2, we have a look from inside the triangle by reviewing the interaction linguistics-biology. Section 3 is devoted to one of the most developed areas in the general framework: bioinspired methods of computation, and their application to natural language processing. Section 4 draws some future possible scenarios for the field that the new science and the new society require. Finally, Sect. 5 presents some general conclusions.

2 Perspectives inside the triangle

In this section, a short overview is provided of the main models that have been introduced on the interaction of

disciplines inside the triangle linguistics-biology-computer science. We focus on the relationship linguistics-biology by means of informatics.

2.1 Explaining the structure of natural language by bio-inspired methods

After the Chomskian revolution, the relationship between natural and formal languages gave rise to a new way to understand language studies. Natural language was thought as an object of the natural world that could be approached from a rationalist science as any other object in the nature. The attempts to describe natural language with mathematics were important milestones in early post-Chomskian linguistics.

Nowadays, linguistic schools prefer the search of new formalisms to account for languages in a simpler and more natural way. Two main facts lead to look for a more natural computational system to give a formal description of natural languages:

1. Natural language sentences cannot be placed in any of the families of the Chomsky hierarchy in which current computational models are mainly based, and
2. Rewriting methods used in a large number of natural language approaches seem to be not very adequate, from a cognitive perspective, to explain how natural language is generated and processed.

Now, if to the above two facts we add that:

3. Languages that have been generated following a molecular computational model are placed in-between context-sensitive and context-free families;
4. The genetic model offers simpler alternatives to the rewriting rules, and;
5. Genetics is a natural informational system as natural language is,

we have the ideal scenario to propose biological models in NLP.

Three main bio-inspired methods are considered in this paper to account for the structure of natural language: DNA Computing (Păun et al. 1998), Membrane Computing (Păun 2000) and Networks of Evolutionary Processors (NEPs) (Castellanos et al. 2003). The contributions of bio-inspired computing to the description and understanding of natural language will be developed in depth in Sect. 3.

2.2 Understanding the genetic code as a linguistic informational system

Since the discovery of the basic structural configuration of DNA (Watson and Crick 1953) many researchers pointed out the communicative nature of the genetic code. This topic, that caused a philosophical discussion for years, has been tackled by authors like Jakobson (1973), Jacob and Monod (1963), Monod (1970), Marcus (1998), Ji (2002), López García (2002), Collado-Vides (1989), Searls (1993) and Victorri (2007). Among them, Jakobson was who introduced the most spread mapping between units in genetic code and their correspondence in natural language.

Some authors, such as Searls (1993) and Brendel and Busse (1984), have successfully used formal languages for the description of DNA. An specially important work for linguistics is the approach by Collado-Vides (1989), Collado-Vides et al. (1998), who was one of the pioneers in the application of methods of analysis taken from generative grammar to explain the functioning of genetic processes of regulation. Perhaps, this is the first example of the exportation of models which are usually used in linguistics to molecular biology.

Formal language theory has also devoted some effort to explain the structure of Genetic Code. Some examples can be the following:

1. Pawlak (1965) dependency grammars as an approach in the study of protein formation.
2. Stochastic context-free grammars for modeling RNA (Sakakibara et al. 1994).
3. Definite clause grammars and cut grammars to investigate gene structure and mutations and rearrangement in it (Searl 1993).
4. Tree-adjoining grammars used for predicting RNA structure of biological data (Uemura et al. 1999).

2.3 Explaining the evolution of natural language as a Darwinian process

In 1953, August Schleicher (1853) introduced one of the most innovative techniques in language evolution: the *use of trees* to illustrate language phylogeny. It is surprising that both, Schleicher and Darwin (1859) came up with the same idea in two distant disciplines: *biology* and *linguistics*. Since that, a lot of effort has been devoted to investigate whether or not some relationship can be established between the phylogenetic trees of both, languages and species.

In the twentieth century, the relationship between biological and linguistic evolution was studied mainly by Cavalli-Sforza (1981). The geneticist has established the relationship between the trees of human and language diversity. Maynard Smith (1997) has also tackled the problem of the relationship between transitions in evolutionist biology and human language, combining his approach with game theory.

Being a key issue in biology and linguistics, traditionally formal language theory has not payed so much attention to the concept of evolution. The word is is synonym for *change by the time* and the main aspects of evolution—*random*

change by replication and *natural selection*—still await to be developed. However, this scenario completely changed in the nineties with the emergence of DNA computing.

Once again, theoretical developments in computer science have a deep impact in other sciences. From the linguistic point of view, the origins and emergence of language are one of the main problems of language evolution, but it was the new computational paradigm that caused the arising of a totally different perspective in research. Nowadays, language evolution has to be studied in collaboration with computer science, biology and artificial intelligence.

One of the first works giving a preliminary overview of the topic is the one by Knight et al. (2000). Some researchers, like Loreto and Steels (2007), are pioneers in the use of simulation for explaining language evolution.

Maynard Smith (1999) considered the emergence of language as one of the major transitions in evolution. The problem of language emergence is crucial in linguistics, but it is also one of the major problems in evolutionary biology, psychology and biolinguistics. The major contributions to the topic of language emergence have been obtained by combining Darwinism, linguistics and computer science. The main exponent of this line of research is Kirby (2001), who introduced the computational model of iterated learning, that could demonstrate the emergence of compositionality despite the lack of physical data.

It would be also interesting to be able to generate and accept languages including the idea of natural selection. The goal would be to formalize the rules of nature in a given environment for generating and erasing symbols in a system. This could help linguistics to approach the topics of language change, language death and language interaction.

This time, it is linguistics the discipline that has forced computer science to fit into the parameters of Darwinism. The main principles of random change by chance and natural selection, that were not developed by computer science, are being conveniently adapted and used for linguists in the simulations of language change and evolution, like in the works by Baxter et al. (2008) and Lieberman et al. (2007).

Finally, it is a current challenge for biology, linguistics and computer science, to be able to formalize the theory of evolution in order to model language evolution. In this way, it would be possible to go from the nature to mathematics and computers and, from here, again to the nature, closing a circle that can help to understand and simulate our amazing linguistic capacity.

Another key issue that biology can help to approach is the emergence and evolution of language. If morphological and cognitive features of humans have to be approached from inside the Darwinian paradigm, it seems clear that the configuration of cognitive and physiological capacities that allow human beings to talk, as well as the process that leads to the arising of human language, have to be tackled from a perspective that integrates biology. However, the lack of data in paleontology and historical linguistics leaves the use of computer science and computational simulations as the only available tool to figure out how the human started to use the symbolic capacity to design a language.

3 Bio-inspired methods for natural language processing

Here, we present an overview of different bio-inspired methods that during the last years have been successfully applied to several NLP issues, from syntax to pragmatics. Those methods are taken mainly from computer science, and can be gathered under the label of molecular computing methods. They are basically the following: *DNA computing, membrane computing* and *networks of evolutionary processors*. The idea is to approach natural language by natural methods by means of the computational formalization of the biological behaviour of living beings.

We want to remark that very important bio-inspired systems of computation, like ant colonies (Dorigo 2004) or swarm intelligence (Beni and Wang 1989) do not fulfill an important requirement for our purposes, namely *formalization*. Ant colonies and swarm intelligence are engineering models and lack the mathematical formalization needed in theoretical linguistics.

On the other hand, the models studied here are all of them placed in a molecular scale. The maximal device that we can construct is a colony or a group of cells collaborating together, never complex mechanisms that cannot be handled by simple mathematical operations. Being this an important feature of such systems, we would like to consolidate *cell computing* to integrate the mathematical formalizations that are needed by theoretical linguistics, all of them based in the smallest level of biological systems.

3.1 DNA computing

One of the most developed lines of research in natural computing is the named molecular computing, a model based on molecular biology, which arose mainly after (Adleman 1994). An active area in molecular computing is DNA computing (Păun et al. 1998) inspired in the way that the DNA performs operations to generate, replicate or change the configuration of the strings.

Application of molecular computing methods to natural language syntax gives rise to molecular syntax (Bel-Enguix and Jiménez-López 2005a). Molecular syntax takes as a model two types of mechanisms used in biology (especially in genetic engineering) in order to modify or generate DNA sequences: *mutations* and *splicing*. Mutations refer to

changes performed in a linguistic string, being this a phrase, sentence or text. Splicing is a process carried out involving two or more linguistic sequences. It is a good framework for approaching syntax, both from the sentential or dialogical perspective.

Methods used by molecular syntax are based on basic genetic processes: cut, paste, delete and move. Combining these elementary rules most of the complex structures of natural language can be obtained, with a high degree of simplicity.

This approach is a test of the generative power of splicing for syntax. It seems, according to the results achieved, that splicing is quite powerful for generating, in a very simple way, most of the patterns of the traditional syntax. Moreover, the new perspectives and results it provides could mean a transformation in the way syntactic mechanisms have been conceived so far.

From here, bio-NLP, applied in a methodological and clear way, is a powerful and simple model that can be very useful to formulate some systems capable of generating the larger part of structures of language, and to define a formalization that can be implemented and may be able to describe and predict the behavior of natural language structures.

3.2 Membrane computing

Membrane systems—introduced in (Păun 2000)—are models of computation inspired by some basic features of biological membranes. They can be viewed as a new paradigm in the field of natural computing based on the functioning of membranes inside the cell. Membrane systems can be used as generative, computing or decidability devices. This new computing model has several intrinsically interesting features such as, for example, the use of multisets and the inherent parallelism in its evolution and the possibility of devising computations which can solve exponential problems in polynomial time.

This framework provides a powerful tool for formalizing any kind of interaction, both among agents and among agents and environment. One of the key ideas of membrane systems is that generation is made by evolution. Therefore, most of evolving systems can be formalized by means of membrane systems.

Linguistic Membrane System (LMS) (Bel-Enguix and Jiménez-López 2005b) aim to model linguistic processes, taking advantage of the flexibility of membrane systems and their suitability for dealing with some fields where contexts are a central part of the theory. LMS can be easily adapted to deal with different aspects of the description and processing of natural languages. The most developed applications of LMS are semantics and dialogue. Membrane systems are a good framework for developing a semantic theory because they are evolving systems by

definition, in the same sense that we take meaning to be a dynamic entity. Moreover, membrane systems provide a model in which contexts, either isolated or interacting, are an important element which is already formalized and can give us the theoretical tools we need. LMS for semantics deal with the main idea that meaning is something dynamic. From that perspective, semantic membranes may be seen as an integrative approach to semantics coming from formal languages, biology and linguistics. Taking into account results obtained in the field of computer science as well as the naturalness and simplicity of the formalism, it seems that the formalization of contexts by means of membranes is a promising area of research for the future.

A topic where context and interaction among agents is essential is the field of dialogue modeling and its applications to the design of effective and user-friendly computer dialogue systems. Taking into account a pragmatic perspective of dialogue and based on speech acts, multi-agent theory and dialogue games, Dialogue Membrane Systems (Bel-Enguix and Jiménez-López 2006) have arisen, as an attempt to compute speech acts by means of membrane systems. Considering membranes as agents, and domains as a personal background and linguistic competence, the application to dialogue is almost natural, and simple from the formal point of view.

3.3 NEPs-networks of evolutionary processors

Networks of Evolutionary Processors (NEPs) are a new computing mechanism directly inspired in the behavior of cell populations. Every cell is described by a set of words (DNA) evolving by mutations, which are represented by operations on these words. At the end of the process, only the cells with correct strings will survive. In spite of the biological inspiration, the architecture of the system is directly related to the Connection Machine (Hillis 1985) and the Logic Flow paradigm (Errico and Jesshope 1994). Moreover, the global framework for the development of NEPs has to be completed with the biological background of DNA computing (Păun et al. 1998), membrane computing (Păun 2000)—that focalizes also in the behavior of cells—and specially with the theory of grammar systems (Csuhaj-Varjú et al. 1994), which share with NEPs the idea of several devices working together and exchanging results.

First precedents of NEPs as generating devices can be found in (Csuhaj-Varjú and Salomaa 1997) and (Csuhaj-Varjú and Mitrana 2000). The topic was introduced in (Castellanos et al. 2003) and (Martín-Vide et al. 2003), and further developed in (Castellanos et al. 2005); (Csuhaj-Varjú et al. 2005). A new approach to networks of evolutionary processors as accepting devices has started in (Margenstern et al. 2004).

With all this background and theoretical connections, it is easy to understand how NEPs can be described as

agential bio-inspired context-sensitive systems. Many disciplines are needed of these types of models that are able to support a biological framework in a collaborative environment. The conjunction of these features allows applying the system to a number of areas, beyond generation and recognition in formal language theory. NLP is one of the fields with a lack of biological models and with a clear suitability for agential approaches.

NEPs have significant intrinsic multi-agent capabilities together with the environmental adaptability that is typical of bio-inspired models. Some of the characteristics of NEPs architecture are the following:

– *Modularization*. NEPs are a distributed system of contributing nodes, each one of them undergoing just one type of operation. Each one of these specialized processors may be modular as well.
– *Contextualization* and redefinition of agent capabilities during the computation.
– *Synchronization*. It is necessary to define protocols for the timing of the computation.
– *Evolvability*. Agents/nodes/modules can change their definition during the computation.
– *Learnability*. The changes in the definition of the agent can be given by the context, by the elements inside or by the learning in their modules.

Inside of the construct, every agent is *autonomous, specialized, context-interactive* and *learning-capable*. Social competences among processors are specified by a graph-supported structure and filtering-regulated communication and coordination.

In what refers to the functioning of NEPs, two main features deserve to be highlighted:

– *Emergence*: The behavior of the nodes is not necessarily pre-established. It depends on both, the evolution of the system and the conditions of the environment.
– *Parallelism*: Different tasks can be performed at the same time by different processors.

Because of those features, NEPs seems to be a quite suitable model for tackling natural languages (Bel-Enguix and Jiménez-López 2005c). One of the main problems of natural language is that it is generated in the brain, and there is an important lack of knowledge of the mental processes the mind undergoes to, finally, bring about a sentence. While expecting new and important advances in neuro-science and neuro-language, we are forced to use the models that seem to fit better to language generation and recognition. Modularity has shown to be a very important idea in a wide range of fields: cognitive science, computer science and, of course, NLP. From this point of view, we think that NEPs provide a quite suitable theoretical framework for the formalization of modularity in NLP.

Another chief problem for the formalization and processing of natural language is its changing nature. Not only words change through the time. Also rules, meaning and even phonemes can take different shapes during the process of computation. Formal models based only in mathematical (an extremely stable) language have a lack of flexibility to describe natural language. Biological models seem to be better to this task, since biological entities share with languages the concept of "evolution" as one of the main features of the system. From this perspective, NEPs offer enough flexibility to model any change at any moment in any part of the system. Besides, as a bio-inspired method of computation, they have the capability of simulating natural evolution in a highly pertinent and specialized way.

Some linguistic disciplines, as pragmatics or semantics, are context-driven areas, where the same utterance has different meanings in different contexts. To model such variation, a system with a good definition of environment is needed. NEPs offer some kind of solution to approach formal semantics and formal pragmatics from a natural computing perspective.

Finally, the multimodal approach to communication, where not just production, but also gestures, vision and supra-segmental features of sounds have to be tackled, refers to a parallel way of processing. NEPs allow to work in the same goal, but in different tasks, to every one of the modules. The autonomy of every one of the processors and the possible miscoordination between them can also give account of several problems of speech.

4 Challenges for the future inside the triangle

The progressive change in the models of computation implies also a change in the problems they can tackle, as well as in the models of cognition they suggest. The first Chomskian languages were based in rewriting, and caused the generalization of the rewriting and tree-like models for describing and explaining natural language. Later on, with the arising of contextual grammars and new typologies of languages not exactly corresponding to the Chomskian hierarchy, new perspectives were started in the formalization of natural language.

A further step in the development of formal languages was the idea of several devices collaborating for achieving a common goal. The new model, besides giving new horizons to formal linguistics, allowed, for the first time, the possibility of approaching with formal methods, not only syntax, but the whole process of language generation.

In the nineties, the emergence of biology as a pilot science had a direct impact in theoretical computer science by means of Molecular Computing, DNA Computing and

Cell Computing. The computability models taken directly from nature, more precisely, from the behavior of DNA replication and communities of cells, provided a natural formalized model for explaining a natural communication system.

New trends in computer science and biology are also shaping the benefits that natural language can take from the interdisciplinary research from inside the triangle. For the future, we think that linguistics will be very much influenced by the answer biology and computer science can give to two challenging tasks:

1. Understanding natural language as complex adaptive system, and
2. Context formalization.

4.1 Natural language as complex adaptive system

The increasing influence of artificial intelligence and the necessity of looking at the systems of the nature with an integrative paradigm have lead to the idea that language can be seen as a complex adaptive system (Holland 1992), a dynamic network (system) where many agents are constantly acting and reacting in parallel, being the system the result of the acts and decisions of individual agents.

Steels (2000) introduces the first analogy natural language-complex adaptive system, that can cause a turnabout in linguistics. First of all, it makes unavoidable the use of interdisciplinarity in the study of natural languages. Secondly, language is understood as a dynamic entity, sharing many features with other complex systems of the nature, namely:

– The system consists on multiple agents that interact each other. In the case of natural languages, this view integrates the behaviour of the speakers, an external factor, in the study of the structure of the system;
– Agents' (speakers') behaviour is usually influenced by social conventions, cognitive or ideological biases, cognitive handicaps or historical contexts;
– The behaviour of the agents is adaptive. It has into account experiences and competition when interacting;
– Language is emergent. It is influenced by past behaviours of the agents, social environment and cognitive processes.

As a result of those features, language must be seen as an evolutionary entity that must be approached from multiple methods, from physics and neuroscience experimentation, to anthropology and history. Probably, one of the best ways to approach this diversity is computational modeling and simulation.

It is known that one of the most relevant complex adaptive systems are living beings. So, biology can again be the model that can help to formalize and understand natural language. On the other hand, this approach has to take into account physics, and the laws of stability and thermodynamics. This means that, still in the realm of the smallest, quantum physics and quantum computing, another increasingly developed field of research can be quite a convenient model for natural language, together with molecular and cell biology.

During the last decade, studies of language emergence and diachrony have undergone a deep transformation thanks to new interdisciplinary approaches introduced from artificial intelligence, physics and evolutionary biology. From the new perspective, language evolution has widely been approached by means of mathematical and computational models involving simulation. Currently, the research in diachronic change is based in the understanding of language as complex adaptive system (Steels 2000) and evolutionary system (Brighton et al. 2005, Croft 2000). All these authors stress the necessity of computational support and simulation to fully understand and describe the processes of adaptation of human communication to the environment.

4.2 Context formalization

The role of context in the comprehension of systemic phenomena is crucial in both, biology and linguistics. Although, it is still in an initial phase.

In biology, the importance of context, referred to the specific terms of environment, ecosystem, cytosol, membrane or system is explicit and widely accepted. Regarding context formalization, like in every area of biology, the help of computation has been important to start the first approaches to the topic.

In recent years, some systems have been developed that allow management and the calculation of biological contexts. We refer mainly to BioAmbients (Regev et al. 2004) and Brain Calculi (Cardelli 2005). These articles offer a system for algebra description of cell membranes and their interactions, mainly applicable to formal biology.

Almost in parallel, Membrane Systems (Păun 2000) were introduced which, based on the behaviour of cells, are formalized as a computational mechanisms. These systems—crucial to understand cellular systems—could inspire powerful computing mechanisms through the formalization of rules that can model the behaviour of membranes of cells.

Membrane systems and BioAmbients differ from their ultimate goal which deviates from the biology, thus giving a new clue that is possible to import models from biological compartments to model other mechanisms.

Natural language, on the contrary, has not had yet a consistent theory for dealing with context. The most useful

contributions come from theories specially developed for informatics. Among them, the approach started by Benerecetti et al. (2001) or the application to context description of Situation Theory by Akman and Surav (1996).

Therefore, context formalization, that has the same key importance in linguistics and computation, must be tackled by biocomputing methods. A first introduction to the area can be given by membrane systems, that seem to be able to provide the theoretical tools that are necessary for the description of context.

However, there are some different computational mechanisms coming from engineering, that can help to build a consistent and cooperative theory of context. Swarm intelligence (Beni and Wang 1989), for example, exploits the possibilities of interaction among agents and environment in a more optimal way.

5 Conclusions

The paper has provided an insight from inside a triangle that has drawn part of the development of the science in the last part of the twentieth century and the first decade of the twenty-first century. Inside the research area delimited by linguistics, computer science and biology, there are still a lot of research possibilities to exploit. Interdisciplinarity is not only a tool, but an essential part of the construct of the nature, that can be seen from many different and complementary perspectives.

Therefore, the full development of networking theories and context theories in bioinspired formal languages, as well as the understanding of natural language as an evolutionary and adaptive complex system can help to tackle some parts of linguistics, like pragmatics, semantics or dialogue, which are crucial for artificial intelligence, and do not have a suitable model to be approached so far.

Developing good models for artificial intelligence is probably the most important task for linguistics and computer science in the future. However, computer science must intervene in the study of language, not only for developing language technologies, but also to improve—by means of the computational simulation of language processes—the research on how language is acquired and processed.

References

Adleman LM (1994) Molecular computation of solutions to combinatorial problems. Sci Agric 226:1021–1024

Akman V, Surav M (1996) The use of situation theory in context modeling. Comput Intell 12(4):1–13

Baxter G, Blythe R, Croft W, Nekane A (2008) Utterance selection model of language change. Phys Rev E 73:046118

Bel-Enguix G, Jiménez-López MD (2005a) Biosyntax. An overview. Fundam Informa 64:1–12

Bel-Enguix G, Jiménez-López MD (2005b) Linguistic membrane systems and applications. In: Ciobanu Gh, Păun Gh, Pérez-Jiménez MJ (eds) Applications of membrane computing, Springer, Berlin, pp 347–388

Bel-Enguix G, Jiménez-López MD (2005c) Analysing sentences with networks of evolutionary processors. In: Mira J, Álvarez JR (eds) Artificial intelligence and knowledge engineering applications: a bioinspired approach. LNCS 3562, Springer, Berlin, pp 102–111

Bel-Enguix G, Jiménez-López MD (2006) Computing dialogues with membranes. Electron Notes Theor Comput Sci 157(4):57–73

Benerecetti M, Bouquet P, Ghidini C (2001) On the dimensions of context dependence: partiality, approximation, and perspective. In: Proceedings of CONTEXT 2001. Springer, Berlin

Beni G, Wang J (1989) Swarm Intelligence in cellular robotic systems. In: Proceedings of NATO advanced workshop on robots and biological systems. Tuscany, 26–30

Brendel V, Busse H (1984) Genome structure described by formal languages. Nucleic Acids Res 12(5):2561–2568

Brighton H, Smith K, Kirby S (2005) Language as an evolutionary system. Phys Life Rev 2:177–226

Cardelli L (2005) Brane calculi, interactions of biological membranes. In: Danos V, Schachter V (eds) Computational methods in systems biology. LNCS 3082, Springer, Berlin

Castellanos J, Leupold P, Mitrana V (2005) Descriptional and computational complexity aspects of hybrid networks of evolutionary processors. Theor Comput Sci 330(2):205–220

Castellanos J, Martín-Vide C, Mitrana V, Sempere JM (2003) Networks of evolutionary processors. Acta Inform 39:517–529

Cavalli-Sforza L (1981) Cultural transmission and evolution. Princenton University Press, Princeton

Collado-Vides J (1989) A transformation-grammar approach to the study of regulation of gene expression. J Theor Biol 136: 403–425

Collado-Vides J, Gutiérrez-Rios RM, Bel-Enguix G (1998) Networks on transcriptional regulation encoded in a grammatical model. BioSystems 47:103–118

Croft W (2000) Explaining language change. An evolutionary approach. Longman, Harlow

Csuhaj-Varjú E, Dassow J, Kelemen J, Păun Gh (1994) Grammar systems. Gordon and Breach, London

Csuhaj-Varjú E, Martín-Vide C, Mitrana V (2005) Hybrid networks of evolutionary processors are computational complete. Acta Inform 41(4-5):257–272

Csuhaj-Varjú E, Mitrana V (2000) Evolutionary systems: a language generating device inspired by evolving communities of cells. Acta Inform 36:913–926

Csuhaj-Varjú E, Salomaa A (1997) Networks of parallel language processors. In: Păun Gh, Salomaa A (eds) New trends in formal languages, LNCS 1218. Springer, Berlin, pp 299–318

Darwin Ch (2009) The origin of species: 150th anniversary edition. Signet classics. New American Library, New York (1st edition: 1859)

Dorigo M, Statzle T (2004) Ant colony optimization. MIT Press, Cambridge

Errico L, Jesshope C (1994) Towards a new architecture for symbolic processing. In: Plander I (ed) Artificial intelligence and information-control systems of robots 94, World Scientific Publisher, Smolenice, pp 31–40

Hillis WD (1985) The connection machine. MIT Press, Cambridge

Holland John H (1992) Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence. MIT Press, Cambridge

Jacob F, Monod J (1963) Genetic repression, allosteric inhibition and cellular differentiation. In: Monod J (ed) Cytodifferentiation and macromolecular synthesis, Academic Press, New York, pp 30–64

Jakobson R (1973) Essais de Linguistique Générale. 2. Rapports Internes et Externes du Language. Les Éditions de Minuit, Paris

Ji S (2002) Microsemiotics of DNA. Semiotica 138(1/4):15–42

Kirby S (2001) Spontaneous evolution of linguistic structure: an iterated learning model of the emergence of regularity and irregularity. IEEE Trans Evol Comput 5(2):102–110

Knight C, Studdert-Kennedy M, Hurford J (eds) (2000) The evolutionary emergence of language: social function and the origins of linguistic form. Cambridge University Press, Cambridge

Lieberman E, Michel JB, Jackson J, Tang T, Nowak MA (2007) Quantifying the evolutionary dynamics of language. Nature 449:713–716

López García A (2002) Fundamentos Genéticos del Lenguaje. Cátedra, Madrid

Loreto V, Steels L (2007) Emergence of language. Nat Phys 3(11):758–760

Marcus S (1998) Language at the crossroad of computation and biology. In: Păun Gh (ed) Computing with bio-molecules, Springer, Singapore, pp 1–35

Margenstern M, Mitrana V, Pérez-Jiménez M (2004) Accepting hybrid networks of evolutionary processors. In: Ferreti C, Mauri G, Zandron C (eds) DNA 10. Preliminary proceedings, University of Milano-Biccoca, Milan, pp 107–117

Martín-Vide C, Mitrana V, Pérez-Jiménez M, Sancho-Caparrini F (2003) Hybrid networks of evolutionary Processors. In: Cantó-Paz E et al. (ed). Genetic and evolutionary computation, GECCO 2003, Genetic and evolutionary computation conference (Part I). LNCS 2723. Springer, Berlin, pp 401–412

Maynard Smith J, Szathmáry E (1997) The major transitions in evolution. Oxford University Press, New York

Maynard Smith J, Szathmáry E (1999) The origins of life: from the birth of life to the origin of language. Oxford University Press, Oxford

Monod J (1970) Le Hasard et la Nécessité. Éditions du Seuil, Paris

Păun Gh (2000) Computing with membranes. J Comput Syst Sci 61:108–143

Păun Gh, Rozenberg G, Salomaa A (1998) DNA computing. New computing paradigms. Springer, Berlin

Pawlak Z (1965) Gramatyka i Matematika. Panstwowe Zakady Wydawnietw Szkolnych, Warzsawa

Regev A, Panina E, Silverman W, Cardelli L, Shapiro E (2004) BioAmbients: an abstraction for biological compartments. Theor Comput Sci 325(1):141–167

Sakakibara Y, Brown M, Underwood R, Saira Mian I, Haussler D (1994) Stochastic context-free grammars for modeling RNA. In: Proceedings of the 27th Hawaii international conference on system sciences. IEEE Computer Society Press, Honolulu, pp 284–283

Schleicher A (1853) Die ersten Spaltungen des indogermanischen Urvolkes. Allgemeine Zeitung fuer Wissenschaft und Literatur

Searls D (1993) The linguistics of DNA. Am Sci 80:579–591

Steels L (2000) Language as a complex adaptive system. LNCS 1917:17–26

Uemura Y, Hasegawa A, Kobayashi S, Yokomori T (1999) Tree adjoining grammars for RNA structure prediction. Theor Comput Sci 210(2):277–303

Victorri B (2007) Analogy between language and biology: a functional approach. Cogn Process 8:11–19

Watson J, Crick F (1953) The structure for deoxyribose nucleic acid. Nature 171:137