



**HAL**  
open science

## Analogy between language and biology - A functional approach

Bernard Victorri

► **To cite this version:**

Bernard Victorri. Analogy between language and biology - A functional approach. *Cognitive Processing*, 2007, 8 (1), pp.11-19. halshs-00137870

**HAL Id: halshs-00137870**

**<https://shs.hal.science/halshs-00137870>**

Submitted on 22 Mar 2007

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## **Analogy between language and biology - A functional approach**

### **Introduction**

Biologists and linguists are used to analogies between their domains of study. The notion of 'Life of Language' is relatively old. It developed in the mid nineteenth century with the advent of Darwin's theory and the foundation of historical linguistics. The dual notion, 'Language of Life' is more recent. It really started with the discovery of the DNA code one century later. But how far can these analogies be pushed and to what extent can they help to understand language and/or life processes? To quote Jacqueline Holzer, talking about the objective of a conference held recently in Sweden, entitled *Genome & Language – Points of Convergence and Divergence*:

“Geneticists, when presenting the structure of the human genome, seem to find the metaphor of the genome as a book, or a text, useful. Genomes and texts are both multiply articulated structures, where purely contrastive units – phonemes, letters, bases – combine to form meaningful units at several levels of increasing complexity – words, sentences, texts; codons, genes, chromosomes.

The question that the symposium tries to answer are the following: Can such points of convergence be pushed further? Can a deeper understanding of form, function, replication, and change of linguistic structures suggest to geneticists new areas of inquiry and new conceptual tools? Can a deeper understanding of form, function, replication, and change of genetic structures help linguists to focus on aspects of language hitherto neglected?” (Holzer, 2002, p. 4).

As a prerequisite to addressing the issue, we must choose on which precise analogy we will focus. Actually, there are several, different analogies that have been put forward in the literature from the beginning, and they need to be carefully distinguished, since they are incompatible with one another. For instance, August Schleicher, one of the founders of historical linguistics, compared human languages with living organisms, claiming that languages are natural organisms that are born, grow, develop, age and die, following definite laws independent of human will (Schleicher, 1863). As for the semanticist Arsene Darmesteter, he related *words* with living organisms, using exactly the same vocabulary. One of his main books (Darmesteter 1886) was entitled *Life of words* and it was composed of three chapters: how words are born, how they live with one another and how they die. On the contrary, Roman Jakobson put words in correspondence with codons (nucleotide triplets), stressing the property of 'double articulation' shared by both genetic and verbal codes:

“Among all the information-carrying systems, the genetic code is the only one which shares with the verbal code a sequential arrangement of discrete subunits – phonemes in language and nucleotides (or 'nuclear letter') in the genetic code – which by themselves are devoid of inherent meaning but serve to build minimal units endowed with their own, intrinsic meaning.” (Jakobson & Waugh, 1987, p.69)

Sereno (1991) classifies analogies between biological and cultural/linguistic evolution into four main groups, according to the nature of the components matched in each domain: species and languages, genes and culture, organisms and concepts, and finally cells and persons. The analogy developed here falls in between the first and the last group, since it does not match a language with a species, strictly speaking, but with the proteome of a species, and it does not match a person with a single cell, but with an organism (the difference vanishes of course in the case of unicellular organisms).

The main originality of our approach consists in taking a functional viewpoint as a starting point. As we shall see, it allows us to give rigorous definitions, building the correspondence on a firm ground. Our approach is close to Pattee's (1980, 1982), Ji's (1999) and Raible's (2001), but with substantial differences that we shall point out when appropriate.

## 1. A functional basis for the analogy

The first function of language is communication and, as such, its fundamental elements are utterances of sentences. The utterance of a sentence is a definite concrete event, the effects of which are observable since it modifies the situation in one way or another, if only in slightly changing hearers' and speaker's states of mind and expectations. It is important to notice that the same sentence may have very different effects depending on the context in which it is uttered. Let us take the example of a very simple sentence. If a person says to her partner "It's too cold" when answering the question "I am going out for a walk, are you coming with me?", she informs her partner that she will not go out because she finds the outside temperature too low (say less than 5°C.). But if the same speaker says to the same partner "It's too cold" while entering a room where a window is open, she probably means that she wants him to close the window because she finds the inside temperature too low (though it could be as high as 20°C. in that case). Therefore the meaning of a sentence cannot be identified with its effects. This is a corollary of the well-known necessary distinction between meaning and reference (*Sinn/Bedeutung*, cf. Frege 1892, Strawson 1971, Ducrot 1984). We must define the meaning of a sentence in a somewhat more abstract manner, as "the constant contribution of the linguistic material composing the sentence to the effects produced by any utterance of the sentence" (Victorri & Fuchs 1996). Carrying on with our example, the meaning of the sentence "It's too cold" has nothing to do with closing a window or going outside or any other action potentially evoked by an utterance of this sentence as a pragmatic 'speech act'. Its semantic content is reduced to the assertion that the temperature of the most salient place in the context is well below the value that could be expected (and desired) by the speaker. This meaning is then incomplete, in a way. It must be combined with information from the contextual situation to deduce what precise range of temperature is asserted for which place.

Turning now to biology, the utterance of a sentence can be compared with the release of a protein (be it intra or extracellular). It is also a definite concrete event which modifies the situation in which it takes place, i.e. the metabolism of the part of the organism in which the protein is released. These events play an essential role in life processes, as essential as language utterances in human communication processes. Effects of the expression of a protein are also context-dependent. It is well known that the modifications induced by the release of an enzyme, for instance, crucially depend on the concentrations of the substrates and often of one or several cofactors as well as on the presence or absence of many potential inhibitors and activators. Many proteins are involved in several different biochemical reactions in different places in the organism. Small protein hormones like insulin and simple polypeptides like angiotensin can have extremely widespread effects throughout the body, acting as neurotransmitters and/or neuromodulators in the brain, as hormones in the blood, as specific catalysts, inhibitors or activators in several types of cells in different target organs (cf. Vincent 1986, pp. 81-82, for a vivid description of the 'polysemy' of these molecules). Interestingly, on the language side of the analogy, very short utterances like "More!" or "Watch it!" can similarly have very different widespread effects, whereas utterances of longer sentences generally act in a much more precise way.

The analogy leads us to define the meaning of a protein as "the constant contribution of the biochemical material composing the protein to the effects produced by any release of the protein". Hence the 'semantic' description of a protein must be made in terms of its biologically relevant chemical properties, such as hydrophobic and hydrophilic domains, number and shape of active sites, ion-channel pores, etc., i.e. all the functional properties that can explain its potential activities in a living organism. The meaning of a protein is of course tightly linked to its 3-dimensional structure, but as is the case for language, meaning must be

distinguished from structure. As we shall see below, the different levels of structure of proteins can be matched, at least partially, with the different levels of the morphosyntactic structure of sentences.

As a consequence of the matching of sentences with proteins, the analogue of a language must be a proteome. Actually, we can define a language as 'the whole set of sentences that can potentially be produced by a speaker' and a proteome as 'the whole set of proteins that can potentially be produced by an organism'. We must notice that these definitions are somewhat abstract since, in both cases, it is not so easy to determine the whole set in question. Before carrying on with the analogy at this level, we must also notice that many previous approaches put forward a correspondence between a language and a genome, instead of a proteome. The reasons why the linguistic notion of double articulation may partly explain such a focusing on the genome will be presented in more detail in the following part of this paper. As for now let us simply say that, from the functional point of view that we have chosen, the genetic material can be considered merely as part of the underlying machinery devoted to producing all the proteins of a given proteome, that is the analogue of the 'organ of language', the computational system postulated by Chomsky which is able to produce all the sentences of a given language.

Two points must be emphasised when matching a proteome and a language. First our definition concerns what was called by Chomsky an 'I-language' (internal or individual language, cf. Chomsky 1986), i.e. the idiolect spoken by a given individual. It must be carefully distinguished from the more common definition of a language as "the totality of utterances that can be made in a speech community" (Bloomfield, 1928), that is what Chomsky calls an E-language (external or extensional language). The analogue of an E-language will be the proteome of a species, i.e. the protein pool obtained by inspecting the proteinic material of every living member of that species. The notion of E-language is even more abstract than its I counterpart since it implies that what a speech community is ought to be determined. The usual definition is based on the notion of intercomprehension (a group of people who are able to understand each other) which is too gradual and vague to provide clearcut partitions. Nevertheless, the notion of E-language can be used as a handy abstraction to schematize an otherwise complex reality. To some extent, the situation is the same in the biological domain where the notion of species also raises some difficulties, interbreeding playing a similar role to intercomprehension in some classical definitions.

The second point concerns the size difference between the set of sentences that can be produced by a speaker and the set of proteins that can be produced by an organism. The first set is often said to be infinite while the other is definitely finite, and in any case much smaller than the first one. The infiniteness of the set of sentences can be discussed. As a matter of fact, as the number of phonemes of a language is finite, the number of sentences is also finite if we assume an upper bound for the number of phonemes per sentence, even if we choose a very large bound, say a million phonemes for instance. For the number of sentences to be infinite, we must support two theoretical claims put forwards by Chomsky and others advocates of generative grammar theories. First we must accept the famous distinction between the 'competence' of the speaker (an idealized capacity) and his 'performance' (his actual production), and we must identify I-language with competence. Then, we must assume that the competence device is a computational system allowing infinite recursivity. Interestingly enough, an analogue of the competence/performance distinction can be found on the biological side, when considering immune system proteins. It is well known (cf. for instance Lefranc & Lefranc 2001), that the number of different immunoglobulins that can theoretically be produced by the human immune system is of the order of billions, whereas the total number of proteins produced in any individual human organism is not thought to exceed a few hundreds of thousands. Here too we see that the size of the human proteome

drastically changes depending on whether we adopt the ‘competence’ or the ‘performance’ point of view.

Nevertheless, beyond the size issue, there is an important point of disanalogy between the two sets. It concerns the productive aspects of the two systems. Even if the postulate of infinite recursivity is rejected, it must be admitted that language possesses productive combinatorial features allowing to generate new sentences very easily. Such general features are missing on the biological side, the immune system being an exception rather than the rule. The differences between both systems resulting from this disanalogy must not be underestimated. As Sereno (1991) playfully says, if the genetic system had the same productive abilities as language, “not only would a giraffe’s neck elongate as it tried to reach the higher branches, but then the animal with the newly acquired long neck would be able to instantly ‘communicate’ such a neck to other short-necked members of the herd simply by ‘lecturing’ to them”. Actually, it is worth noting that the immune system is able to react with a flexibility and a speed that come closer to the inventiveness of language. An interesting though very limited analogue of communication is even to be noticed: the transfer of antibodies from mother to offspring. Generally speaking, there is a closer proximity between the immune system and language which should be interesting to further investigate.

## 2. Hierarchical structures

If we now turn to the structural aspect of the analogy, the first observation to be made is that in both cases there is a primary sequential structure forming the basis of a complex hierarchical organisation. As regards proteins, the discrete units composing the sequence are the twenty proteinogenic amino acids composing the polypeptide chain. As for language, the discrete units are the phonemes. Their number changes from one language to another, but the order of magnitude remains the same as the number of amino acids. The size of the sequence varies widely in both cases: from a few units to thousands for sentences and from a few tens (if we include small hormone polypeptides) to tens of thousands for proteins.

At the first level of organisation in sentences we find words, or, more precisely, morphemes. Morphemes are the smallest units conveying meaning, such as lexical roots, prefixes, suffixes, grammatical inflections, and function words (determiners, prepositions, conjunctions and so on). They consist in strings of adjacent phonemes and the entire chain can be segmented in a sequence of adjacent morphemes. Morphemes are the basic units of the upper levels of organisation. In other words, the phonemic composition of a morpheme has no direct influence on the way it interacts with the other morphemes composing the sentence. This is the property of language known as ‘double articulation’.

Morphemes are organized in hierarchically structured phrases, which in turn form possibly hierarchically structured clauses composing the sentence. The overall organisation of the sentence is generally represented as a syntactic tree whose leaves are morphemes. A tree representation is more accurate than a sequential one-dimensional one because syntactic relations do not necessarily concern adjacent phrases (the phenomenon is called ‘long-distance dependency’ by linguists). Some of them can even link elements which do not belong to the same clause in the ‘surface form’ of the sentence, as shown by the following example, where *who* is the direct object of the verb *invite*:

*Who did you say that Mary knows that Alice did not invite to the party?*

Thus, the global syntactic structure of a sentence cannot be directly derived from its sequential structure. It depends on specific properties of the morphemes present in the sequence. For instance, different verbs do not need the same number of arguments, and most of them impose additional constraints on the very nature of these arguments.

To determine the meaning of the sentence, one must rely on the syntactic structure, rather than on the sequential one. Different word orders can lead to approximately the same meaning

if they share the same syntactic tree, as shown by the sentences: *John gave Mary a present* and *John gave a present to Mary*. Conversely, the same word order can hide different meanings if different syntactic trees can be built from the same surface form, leading to an ambiguity, as in the sentence *John saw a man with binoculars*, where the phrase *with binoculars* can be linked to the noun phrase *a man* or the verb *saw*.

Turning to proteins, we can observe a similar hierarchical organisation, with the secondary, tertiary and quaternary structures. Besides we have the same intimate relation between the 'meaning' of a protein (defined as its biologically relevant chemical properties, cf. above) and its global structure, i.e. its three-dimensional configuration. We can even have the same structural 'ambiguity' of the primary sequential structure since some polypeptides can have more than one stable folded conformation, each conformation having its own biological activity. Protein folding depends on links between different parts of the sequence, and here too, long-distance interactions are common. It is also worth noting that the specific processes (like the action of 'chaperone' proteins) which control the proper folding of some proteins can be considered as the analogue of prosody in oral language which helps in finding the right syntactic structure of some sentences.

There is yet an important disanalogy between the two hierarchical structures (cf. table 1 for an overview of the correspondence). Nothing in proteins can match the first level of organisation of a sentence, that is the grouping of phonemes into morphemes. At the first level of organisation for a protein, we find the alpha-helices and beta-sheets of the secondary structure, which are actually groupings of adjacent amino-acids. But they cannot be compared with morphemes (nor words, for that matter) because they do not form a complete and systematic segmentation of the chain. More importantly, it cannot be said that the detailed amino-acid composition of an alpha-helix or a beta-sheet has no influence on its interactions with other parts of the molecule, as it is the case for the phonemic composition of a morpheme. In other words, there is no double articulation on the biological side of the analogy that we propose here.

This claim may seem somewhat disappointing, as double articulation has been presented as an outstanding feature in almost all the previous approaches of the analogy since the discovery of the DNA code, starting with the great linguist Roman Jakobson himself, as we already mentioned in the introduction. As a matter of fact, the double articulation property put forward by Jakobson does not concern proteins, but DNA itself: phonemes are matched with nucleotides, and morphemes with codons. There is unquestionably a double articulation at this level: codons are the minimal units involved in gene expression, the nucleotide chain can be segmented in codons (putting aside the issue of non-coding DNA, which can be properly tackled in this context), and nucleotides only intervene by their belonging to a triplet. But the analogy falls short as far as semantic and, above all, functional aspects are concerned. The matching of codons with words entails that genes match sentences. Hence the 'meaning' of a gene must be its own contribution to the biological processes in which it is involved, that is to say the synthesis of messenger RNA (mRNA). At best, if we identify in some way a gene both with the portion of DNA and the corresponding mRNA, the meaning of a gene may also be the translation process by ribosomes. This is not fully satisfying since these events are only a small, though not anecdotic, part of what happens in a cell. Here, the whole organism's metabolism can only be taken into account as a far side effect, contrasting with the central place it has in our approach. Moreover, concerning the structural aspects, all the hierarchical complexity is lost: there is nothing in gene structure corresponding to the secondary, tertiary and quaternary structures of proteins.

Many authors (Pattee 1980, Sereno 1991, Raible 2001) tried to save both the interesting features of proteins and the double articulation property by combining in a same complex analogy DNA and protein components. It seems more reasonable to simply give up the double

articulation property. Actually, the DNA code looks like the Morse code more than a real language. The fascination of these authors for the double articulation feature does not seem justified. Its main virtue in language is that it considerably increases the combinatorial power of the system, jumping from a few tens of phonemes to several tens of thousands morphemes. As we already noticed, combinatorial power is precisely what makes the most clear cut distinction between the two systems.

More interestingly, Ji (1999, 2006) considers ‘DNese’, ‘RNese’, and ‘proteinese’ as distinct sub-languages of the cell language, that he calls ‘cellese’. He develops separate analogies between ‘humanese’ and each of these sub-languages. In the analogy involving proteins, he matches morphemes with functional domains (secondary structure), words with folded polypeptides (tertiary structure), and sentences with protein complexes (quaternary structure). He then endows proteinese with the property of double articulation, on the basis of this parallel (he even speaks of a “triple articulation”, taking into account the level of texts: we will come back to this point further down). It seems difficult to fully accept Ji’s claim, because the combinatorial aspect is generally missing. Once more, the only domain where the claim is entirely justified is the immune system, even if, as far as I know, Ji himself does not mention it. Actually, immunoglobulins can be considered as sentences, the heavy and light chains as phrases (rather than words), and the different domains composing the chains as morphemes. We may even identify the constant domains with grammatical morphemes (they play a more structural role), and the variable ones with lexical items (more semantic role). But if we don’t want to restrict the analogy to the immune system alone, it seems more interesting to consider the quaternary structure as an equivalent of the arrangement of clauses inside a sentence. Many proteins are composed by only one polypeptide chain, just as many sentences are made of only one clause, and in protein complexes, the polypeptide chains can be more or less tightly bound, just as clauses of a same sentence.

Above the sentence, following Ji’s ideas, we can carry on the analogy to the upper level, identifying speeches (or texts, as their written equivalent) with what he calls “protein networks”. More precisely, we define a speech as an ordered set of utterances forming a coherent unit in communication. The size of the set may vary widely, from the utterance of a single sentence to several tens or more. With this definition, a very simple sentence like *Watch it!* can be a speech by its own, as well as a long monologue. The parallel with proteins is then clear. Some proteins may be active by their own, but often the functional efficiency is due to a set of proteins acting together. For instance, the Krebs cycle is controlled by a whole set of enzymes acting coherently for that purpose. The speech level is much less precisely determined than the sentence level. It is often difficult to decide if a topic change in a speech is drastic enough to delimitate two distinct ‘speeches’ in the sense we are using the word here. But on the other hand, the speech level is unavoidable since phenomena like anaphora can extend over a whole speech, making the meaning of some sentences unintelligible if considered in isolation. The same can be said of cell metabolism, where biochemical reactions form a network which is difficult to delimit, whereas the action of many proteins is difficult to understand without considering as a whole a large part of the network.

	<b>Linguistics</b>	<b>Biology</b>
<b>Functional events</b>	utterance of a sentence	release of a protein
<b>Definition of meaning</b>	the constant contribution of the linguistic material composing the sentence to the effects produced by any utterance of the sentence	the constant contribution of the biochemical material composing the protein to the effects produced by any release of the protein
<b>Basic structure</b>	chain of phonemes (several tens of different phonemes)	chain of amino-acids (twenty different amino-acids)
<b>Hierarchical levels</b>	<ul style="list-style-type: none"> <li>- phonemes (double articulation)</li> <li>- morphemes</li> <li>- phrases</li> <li>- sentences</li> <li>- speeches</li> </ul>	<ul style="list-style-type: none"> <li>- amino-acids (primary structure)</li> <li>- alpha-helices, beta-sheets (secondary structure)</li> <li>- folded polypeptides (tertiary structure)</li> <li>- protein complexes (quaternary structure)</li> <li>- protein networks</li> </ul>
<b>Meaningful structure</b>	Tree-like syntactic structure of the sentence	Three-dimensional structure of the protein
<b>Individual system</b>	I-language of a speaker	proteome of an organism
<b>Collective system</b>	E-language of a speech community	proteome of a species
<b>Slow dynamics</b>	Language change	Species evolution

**Table 1.** An overview of the analogy



### 3. Evolution of languages and evolution of species

If we now look at evolutionary aspects, we can observe that the analogy still works to a certain extent. It has long been noted that languages differentiate via a process similar to the biological process of speciation. Linguistics communities separate off one from the other (typically geographically). In each of the new communities, the language changes differently. And in time people from separate communities no longer understand each other. Two new languages are born, and they are both descendants of the previous language. Relationships between languages sharing a same ancestor are also called 'genetic' relationships, and they are classically represented as a tree diagram.

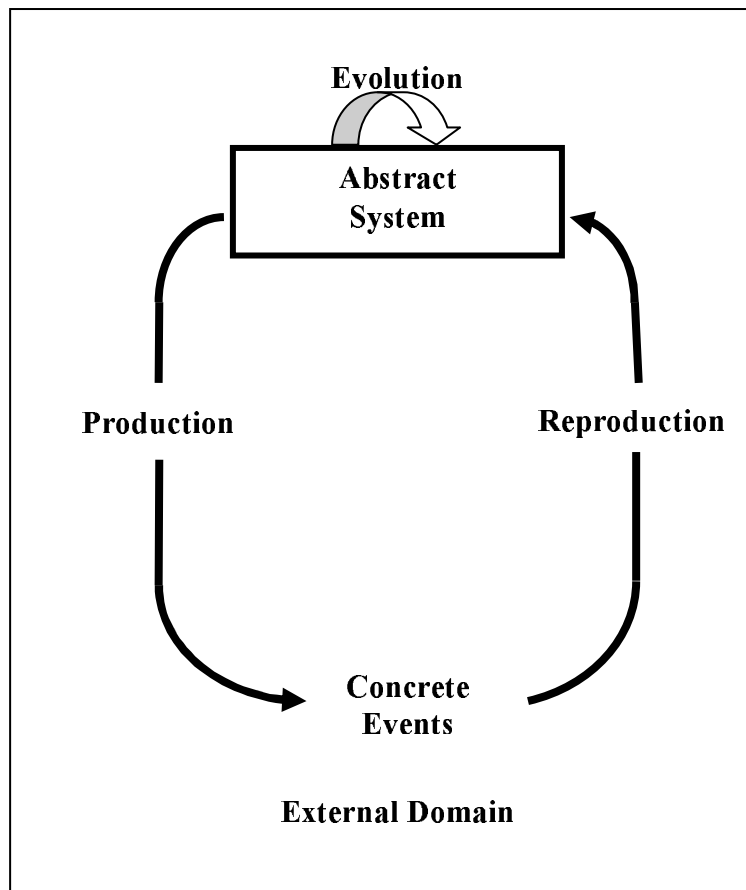
It is worth reminding the distinction between E-language and I-language in this context. When we say that languages change, we are speaking of E-languages, that is to say languages shared by a community of speakers. But of course, there cannot be any change in a E-language if it did not first appear in a I-language and if it did not spread through a significant part of the community's I-languages. The situation is similar in biology. Any change in the proteome of a species (which is, as we saw above, not easier to define rigorously than an E-language) must have appeared and spread at the individual level.

What seems very different between biology and language is the exact mechanism of change. Changes in proteome are a consequence of random genetic mutations in DNA submitted to natural selection pressure. They take place during the reproduction process and their diffusion entirely depends on the reproductive success of the individuals possessing the new features. On the contrary, language innovations are cultural phenomena, that can be produced at any time in a speaker's life and that can be adopted by a large number of members of his community almost immediately. But taking a closer look at linguistic change proper, we can observe that the difference is not so great. First, most often than not, linguistic changes do not result from conscious deliberate decisions, be they individual or collective. Speakers' innovations are constant and numerous in everyday life, but most of them are fleeting trials, lost as soon as they have been uttered. Only a few of them survive and eventually succeed in being integrated in the language standards. In other words, there is a form of selection among a wide range of potential variations, as in biology, even though those modifications are not produced randomly, strictly speaking. Moreover, for a change to become permanent, to spread widely in a social group is not enough: it could only be a fashion phenomenon which is likely to disappear as fast as it came. Therefore, passing from one generation to the next is a necessary step in the process. In other words, language acquisition by children plays an essential role in language change which can be compared, at least partially, with the role of reproduction in species evolution.

We can summarize the evolutionary aspects of our analogy as follows. Languages and proteomes are what can be called 'productive self-reproductive systems' (cf. figure 1).

- the system is productive, in the sense that it is a generative component which enables the production of significant events in an external domain, i.e. events playing an essential part in the dynamics of the processes occurring in the external domain. For languages, the external domain is language communication in a community (*la parole*, in Saussurian terms, while the system itself is *la langue*, cf. Saussure 1972), and the events are utterances. For proteomes, the external domain is the metabolism of an organism, and the events are release of proteins.

- the system is self-reproductive, in the sense that it is maintained and renewed through the events produced in the external domain. The important point is that the system cannot survive in the absence of those events. This is of course true for proteomes which need organism reproduction for their survival, as well as for languages, which irremediably die when they are no more acquired by children as their mother tongue.



**Figure 1.** Productive self-reproductive system

Two dynamics are at work in those systems: a fast dynamics in the external domain, every event somewhat modifying the situation, and a slow dynamics in the system itself, due to the interactions with the external dynamics through the reproduction process. Thus such systems are evolutionary by nature, whatever the details of the mechanism governing the slow dynamics.

There are several other types of productive self-reproductive systems in the cultural sphere. Let us briefly mention two of them, legal systems and scientific systems. In the legal domain, the system is the law, defined as a set of rules or norms of conduct, the external domain is the whole set of legal institutions, and the events are judgments. In the scientific domain, the system is a theory in a given scientific field, the external domain is the whole set of scientific institutions working in the field, and the events are experiments.

We can learn much by comparing the different types of productive self-reproductive systems. For instance, it can be observed that the evolutionary dynamics of all the systems have the same characteristics, although they greatly vary in their mechanisms: periods of small gradual local changes alternate with moments of more abrupt and global reorganizations. This observation suggests that it could be most interesting to study evolutionary dynamics at a more abstract level, in the framework of a general theory of productive self-reproductive systems.

We shall not go further on this issue, but focus on only one topic in which languages show a specificity that the other systems, be they from the biological or the cultural sphere, do not share. Coming back to the issue at stake in this paper, this specificity will be interpreted as an important limit of our analogy. The key question is not *how* the system evolves but *why*. In all the other systems, adaptation seems to be the right answer. In biology, species evolution

obeys a principle of optimality: each species evolves so as to be better adapted to occupy its ecological niche. In the legal domain, laws evolve so as to be better adapted to changing needs in the regulation of social interactions. In the scientific fields, evolution of theories also aims at better taking into account experimental facts. Strangely enough, language change is not driven by a similar optimization principle (cf. for instance Lass 1997, Lightfoot 1999). Actually every known language possesses the same power of expression, and it cannot be said that English or French, for instance, are better adapted to human communication than dead languages like Latin or Ancient Greek. On the long time scale, the evolution of languages appears to be neutral, rather than adaptive.

The reason seems to be that adaptation can occur *inside* a language, and so, it is not necessary to change a language to adapt to new communication needs. As we saw, the combinatorial properties of languages enables the creation of new words and new expressions to fulfil the need to talk about new things and new situations, as illustrated by Sereno's example of the 'acquired' and 'lectured' giraffe's neck mentioned above. Once more, we find that an important disanalogy between language and biology is due to the difference of generative power between the two systems.

### **Conclusion**

Are new perspectives opened by our analogy? How can we answer Holzer's questions mentioned at the beginning of this paper, i.e. can the analogy suggest to geneticists new areas of inquiry and new conceptual tools, and help linguists to focus on aspects of language hitherto neglected?

Concerning linguistics, it seems worthwhile to mention the last problem that we have had on face. Why do languages continue to evolve, if no optimization is at stake? Historical linguistics has focused on the different mechanisms producing language change, such as phonetic erosion (cf., among many others, Bybee 2001), grammaticalization (Hopper & Traugott 1993), and so on. But these studies intend to answer the how question, not the why. Sociolinguistic phenomena (cf. Labov 2000) may to a large extent explain language change. The creation of new linguistic forms or creative adaptation of old ones are often initiated by social groups of all sorts seeking to assert their identity by using their own particular variety of the common language. Pure psychological aspects of language acquisition would also be interesting to study in that perspective. Language acquisition is an important component of the construction of the personality of the child. Thus children also need to appropriate their own language to assert their identity.

As for biology, there is at least one domain, immunology, in which the analogy could suggest new conceptual tools. As we saw, the immune system seems to be the closest biological equivalent to language. John Stewart (cf. Stewart & Varela 1994, Stewart & Coutinho 2004) has developed an interesting account of the 'cognitive properties' of the immune system in the framework of Francisco Varela's autopoiesis theory. At first sight, it seems to be the best framework in which our analogy could be pursued.

Lastly, the analogy led us to define a general class of systems, the productive self-reproductive systems. This is certainly the most important kind of result that can actually be expected from an analogy. Since an analogy is not, by definition, a perfect match between two phenomena, it is natural, determining their common features, to seek for other phenomena sharing these features. The class of phenomena that is then obtained is often the right level at which to study important aspects underlying the analogy. Hopefully, in our case, the class of productive self-reproductive systems could help deepen our understanding of evolutionary dynamics.

## References

- Bloomfield L., 1928, A Set of Postulates for the Science of Language, *Language* 2.
- Bybee J., 2001, *Phonology and Language Use*, Cambridge University Press, Cambridge.
- Chomsky N., 1986, *Barriers*. MIT Press, Cambridge.
- Darmesteter A., 1886, *La vie des mots étudiée dans leurs significations*, Delagrave, Paris.
- Ducrot O., 1984, *Le dire et le dit*, Paris, Editions de Minuit.
- Frege G., 1892, Über Sinn und Bedeutung, *Zeitschrift für Philosophie und philosophische Kritik*, 100:25-50.
- Holzer J., 2002, Concluding reflections, *Colloquium "Genome & Language – Points of Convergence and Divergence"*, Birgitta Forum, Vadstena, Sweden.
- Hopper P. J., Traugott, E. C., 1993, *Grammaticalization*. Cambridge University Press, Cambridge.
- Jakobson R., Waugh L. R., 1987, *The Sound Shape of Language*, de Gruyter, Berlin, New York.
- Ji S., 1999, The Linguistics of DNA: Words, Sentences, Grammar, Phonetics, and Semantics, *Annals of the New York Academy of Sciences*, 870:411-417.
- Ji S., 2006, The proteome as a molecular language (proteinese), *DIMACS Workshop on Sequence, Structure and Systems Approaches to Predict Protein Function*, Rutgers University, New Jersey.
- Labov W., 2000, *Principles of Linguistic change. Volume II: Social Factors*, Blackwell, Oxford.
- Lass R., 1997, *Historical Linguistics and Language Change*, Cambridge University Press, Cambridge.
- Lefranc M.-P., Lefranc G., 2001, *The Immunoglobulin FactsBook*, Academic Press, London.
- Lightfoot, D., 1999, *The development of language: Acquisition, change and evolution*, Blackwell, Oxford.
- Pattee, H.H., 1980, Clues from molecular symbol systems, In Bellugi/Studdart-Kennedy, *Signed and Spoken Language*, 261-274.
- Pattee, H.H., 1982, Cell psychology: An evolutionary approach to the symbol-matter problem, *Cognition and Brain Theory*, 5: 325-341.
- Raible W., 2001, Linguistics and Genetics: Systematic Parallels, In Haspelmath M., König E., Oesterreicher W., Raible W. (eds), *Language Typology and Language Universals*, de Gruyter, Berlin, New York.
- Saussure F. de, 1972, *Cours de linguistique générale*, Payot, Paris.
- Schleicher A., 1863, *Die Darwinsche Theorie und die Sprachwissenschaft - offenes Sendschreiben an Herrn Dr. Ernst Haeckel.*, H. Boehlau, Weimar.
- Sereno M.I., 1991, Four analogies between biological and cultural/linguistic evolution, *Journal of Theoretical Biology*, 151:467:507.
- Stewart J., Coutinho A., 2004, The Affirmation of Self: A New Perspective on the Immune System, *Artificial Life*, 10:3, 261-276.
- Stewart J., Varela F.J., 1994, L'intelligence collective des lymphocytes, in E. Bonabeau (ed.), *L'Intelligence Collective*, Hermes, Paris, 145-156.
- Strawson P.F., 1971, *Logico-linguistics papers*, Ashgate Pub., Aldershot, Burlington.
- Victorri B., Fuchs C., 1996, *La polysémie – Construction dynamique du sens*, Hermès, Paris.
- Vincent J.-D., 1986, *Biologie des passions*, Odile Jacob, Paris.