**General Linguistics and the Theory of Communication**
Palacký University · Olomouc

**Philosophy, Science, Cognition, and Semiotics**
Alma Mater Studiorum · Università di Bologna

# A Linguistic Approach to Protein Folding

## Towards a Semiotic Description of Living Systems

ĽUDMILA LACKOVÁ

**Supervisors**

MGR. DAN FALTÝNEK, PH.D.
Palacký University · Dept. of General Linguistics, Faculty of Arts

PROF. CLAUDIO PAOLUCCI
Università di Bologna · Dept. of Philosophy and Communication Studies

Ph.D. Thesis | 2018

# Declaration

I declare that this dissertation is the result of my individual and independent research. All of the sources consulted have been properly cited.

Olomouc, August 29th 2018

…………………………………………………………………

Signature

# Acknowledgements

First and foremost I want to express my deepest gratitude to my supervisors Mgr. Dan Faltýnek, Ph.d. and prof. Claudio Paolucci for the guidance and abundant inspiration.

I'd like to thank my colleagues and friends for discussing with me the ideas presented in this thesis. I am very grateful in particular to Diego Krivochen not only for his numerous and precious comments and help with this thesis, but also for all years of inspiring conversations and collaboration, encouragement and insightful feedback.

Thanks also to UPLIFT language school for the language corrections and comments.

I wish to thank my family, and especially to my parents and my siblings for their support and for their love throughout these years. The most part of this thesis has been written in presence of my mum in a creative environment of a shared office, coffee and wine. Thanks, mum, for sharing with me those moments.

My utmost thanks and sincere gratitude to all my mentors from Palacký University. Thanks to prof. Jan Kořenský, my mentor for over five years, whose work provided me a major inspiration for this thesis. Thanks to Mgr. Lukáš Zámečník, Ph.D. for fruitful exchange of ideas that contributed to some parts of this thesis and also for plenty of support.

A special and separate ‘thank you’ goes to Marco Petolicchio who has witnessed and contributed to the development and concretization of this work and supported me- especially in the very last weeks of writing. Without him, this work would never come to the present form.

# Contents

# List of Tables

# List of Figures

# 1

# Introduction

This dissertation will try to answer a seemingly simple question: How to construe the grammar of a language? "Language" here is understood in the broad sense, in other words, the term "language" is not limited to natural language phenomena.

As Bolshoy pointed out, there is a

> wide expansion of the number of objects that this term is applied to: language of bees, dolphins' language, programming language, queries language, DNA language, etc. (Bolshoy et al. 2010).

However, I am aware of the fact that one cannot freely rank natural language with gibbon calls or computer programming language. While the former is believed to lack syntax (Fischer 2018), the latter one is believed, on the contrary, to operate only syntactically (Searle 1980). Without denying the uniqueness of natural language, I believe that some analogies and comparisons with non-linguistic sign systems

Figure 1.1: Man Ray (Emmanuel Radnitzky). Fortune. 1952. Wood, plastic beads, and felt, 9 x 28 x 18 cm. The Vera and Arturo Schwarz Collection of Dada and Surrealist Art in the Israel Museum, Jerusalem, Israel. Source: http://www.imj.org.il/en/collections

might bring new light to the understanding of natural language and vice versa; the natural language metaphor may help to understand other sign systems. The analogy or comparison has to be based on a relevant item, in other words, the very object of comparison must be chosen to bring potential relevant results. Charles F. Hockett wrote in his famous paper “The Origin of Speech”:

> The useful items for this sort of comparison cannot be things such as the word sky; languages have such words, but gibbon calls do not involve words at all. Nor can they be even the signal for danger, which gibbons do have. Rather, they must be the basic features of design that can be present or absent in any communicative system, whether it be a communicative system of humans, of animals or of machines. (Hockett 1960, p. 5)

In the following pages of his paper, Hockett enumerates and comments on thirteen such features of the design of languages, among them being discreteness, arbitrariness, semanticity, productivity, displacement, and so on. His theory is based on the similarity between animal communication and human language. The focus of my dissertation will be, on the contrary, on protein folding, a fascinating biological process currently under scientific investigation. Still, Hockett's argument is relevant also for my thesis, since I am focusing on a very basic feature of the design of language (whether human language or "protein language"), which is even more basic than Hockett's thirteen features. I propose that the analogy between natural language and the protein folding process is linked by only one special design feature. This feature can be termed as the asymmetrical relation between linearity and non-linearity. Both grammars of natural language and protein studies deal with linearity problems, that is to say, they try to resolve the tension between linearity on one hand, and non-linearity on the other. This is seen in the tension between sound and meaning in natural language and the tension between the peptide chain and protein shape in protein folding.

It can be argued that the analogy has no scientific value; still it is used in science rather frequently, especially in biology. The whole theory of evolution is based on finding analogies. In addition, cross-field analogies are not rare (for example, in the case of DNA language). But the use of analogy is not exclusive in the field of biology. C. S. Peirce used an analogy with chemical diagrams to represent logical relative terms and algebraic invariants. He was convinced that this enables us to understand properties of algebra which would otherwise escape from us. He said:

> The proposition "John gives John to John" corresponds in its constitution [...] precisely to ammonia. But beyond this point the analogy ceases to be striking. In fact, the analogy with the ruling theory of chemical compounds quite breaks down. Yet I cannot resist the temptation to pursue it. After all, any analogy, however fanciful, which serves to focus attention upon matters which might otherwise escape observation is valuable. (Peirce CP, § 3.470)

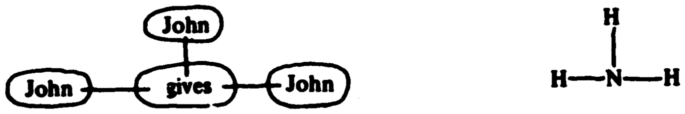

Fig. 5 Fig. 6

Figure 1.2: A chemical atom is quite like a relative. (Peirce CP, § 3.469)

Besides that analogy serves to focus attention upon matters which might otherwise escape observation, it may lead the research to interesting results. A fanciful analogy between protein folding and natural language ceases to be an empty analogy in the moment when one realizes that the asymmetric relations between linearity and non-linearity create the elementary structure, a constitutive principle that both proteins and language have in common. I name this structure a linguistic structure, but the name itself is of little importance.

# 2

# Protein folding

## 2.1 Proteins are folded strings

Proteins, the smallest functional units of our bodies, are complex organic macromolecules, and probably existed at the very beginning of life of Earth[1]. All discussions about the origins of life or about the origin of the genetic code are intrinsic to the discussion about the origin of proteins (Crick 1968). In fact, the very reason for the existence of the genetic code rests on the need for the synthesis of proteins. Our cells use the genetic code to preserve texts with information about which proteins to produce and how to produce them. The miracle of

---

[1] The origin of life on Earth is unknow, yet the most plausible hypotheses, so far, are of two basic types: it was nucleic acids first (code) or it was proteins first (what the code codes for). The supporters of protein hyposthesis are not rare, e.g. (Andras and András 2005). Besides code and proteins, there is also a third option, the so called "RNA world hypothesis" (Robertson and Joyce 2010).

life is hidden in protein synthesis. Proteins are the elementary building blocks of life; they constitute cells and take care of all metabolic processes.

Proteins are nothing but folded strings of hundreds of amino acids. An unfolded string is just a long chain of molecules, yet the moment in which it folds and deflects into a compact bunch, miraculously a three dimensional and functional structure, which guides all organic function in our bodies, emerges. Proteins have been an interesting topic for philosophical essays since the very beginning of modern molecular biology and genetics (Jacob 1970; Monod 1972; Deleuze and Guattari 1987): the process of change from the linearity of the original peptide chain to the dimensionality of the final protein has been especially stimulating for fruitful philosophical discussions.

In the last few years, the quest for protein folding is being in the centre of scientific inquiry more than ever before in the history of biology. Thanks to progress in informatics and newly created fields such as proteomics or bioinformatics, scientists have strong tools to quantitatively analyse a huge amount of data and to try to predict protein structure exclusively from the sequence of amino acids within a peptide chain. However, we do not know how exactly amino acid chains turns into a three-dimensional protein structure. Protein folding, a process of getting to the final protein structure of the original peptide chain, is a very obscure process thus far, and the rules governing the protein folding are waiting to be discovered. Perfectly knowledge of the chemical and physical properties of the peptide chain with the potential to fold does not help that much in understanding the question of why proteins fold in that way and not another. Chemistry and physics play a role yet don’t seem to be the crucial answer. This requires an explanation of another kind: “protein code”, “protein grammar” and “protein syntax” are terms which occur because “a code” is being sought rather than a purely chemical explanation. This situation is very similar to the deciphering of the genetic code, for as much as it is believed that protein folding is a mechanism that was obtained by natural selection, meaning that it was achieved in a way similar to the genetic code by evolutionary convention rather than chemical or physical necessity.

It can be said that molecular biology, and proteomics in particular, is a somehow related to structuralism (in the sense of linguistic structuralism from the last century, that held to the favour of form over substance). Proteins have their metabolic functionality thanks to their shapes or spatial structures, a material realisation being of minor importance. In the same way, languages are—in terms of structural linguistics—structures of relationships between elements, completely dissociable from their material realisation. To put it in another way, structural linguistic supposes that there is a set of abstract relations which do not require their material realization. To give an illustration, consider interlingual translation, where an element can be successfully translated across various languages; thus, the same linguistic meaning can be mediated in different languages thanks to independence of the form of the linguistic substance. The same principle is valid for a protein and its function. The function of proteins is recognised exclusively by its structure. By virtue of this property, in the 1970s researchers began to chemically manufacture artificial proteins that have a completely different chemical design than natural proteins, yet the same structural relations. Peptoids, or synthetic proteins, have the potential to be used for pharmaceutical purposes.

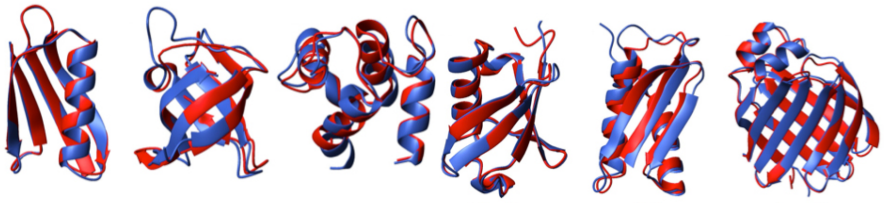

Figure 2.1: Protein structures. Source: https://www.researchgate.net/Backbone-ribbon-representations-32-of-the-lowest-energy-CS-ROSETTA-structure-red_fig3_5524871. Accessed 16 Aug, 2018

This recalls Saussure's famous statement that "language is a form, not a substance" (de Saussure 2011, p. 122). The linguistic analogy in biology has never been as striking as in case of protein folding. Before I go forward with my argument, I propose, for the sake of clarity, to introduce some basic biological principles.

## 2.2 Protein synthesis and proteins. Basic principles.

In the process of protein synthesis, the DNA sequence is transcribed into the RNA sequence (transcription), which is in its turn translated into amino acids (translation). Even though approximately 500[2] different types of amino acids occur in nature, only twenty of them were selected by natural selection to play role in protein synthesis. After the translation process, the amino acids bind together so to form a chain, a sequence called peptide chain. This process is a kind of manufacturing of an amino acid chain. The amino acid chain provides the material for the protein, which is consequently built up from the amino acid chain simply by its folding. Not all twenty amino acids are necessary to construct a protein; it was proven experimentally with artificially synthetized proteins that a small structure may be build up from a peptide chain composed of combinations of only three types of amino acids (Berezovsky, Guarnera, and Zheng 2017).

In agreement with traditional views on protein folding, the folding process has three steps. Firstly the primary structure (the peptide chain) is folded to form secondary structures[3]. Secondary structures are consequently bonded together to form the tertiary structure. Afterwards, it is possible to obtain the quaternary structure by combining two or more tertiary structures together.

Now let's have a further look to the amino acids. The amino acid is composed of amine, carboxylic acid and the residue (R group). Amino acid residues are important because they are the special part which gives the amino acid its uniqueness. Amino acids bond together form

---

[2] See (Cole and Kramer 2016), it should be noted however that this number is only approximate, in fact, there are some specific amino acids present in mitochondria (King 2007).

[3] The importance of secondary structures in protein folding has been questioned by Berezovsky and Trifonov (Trifonov et al. 2001; Berezovsky, Guarnera, and Zheng 2017) who proposed to consider the "closed loops" as having more importance in the very process of folding, secondary structures only playing role in the final detailing of a protein.

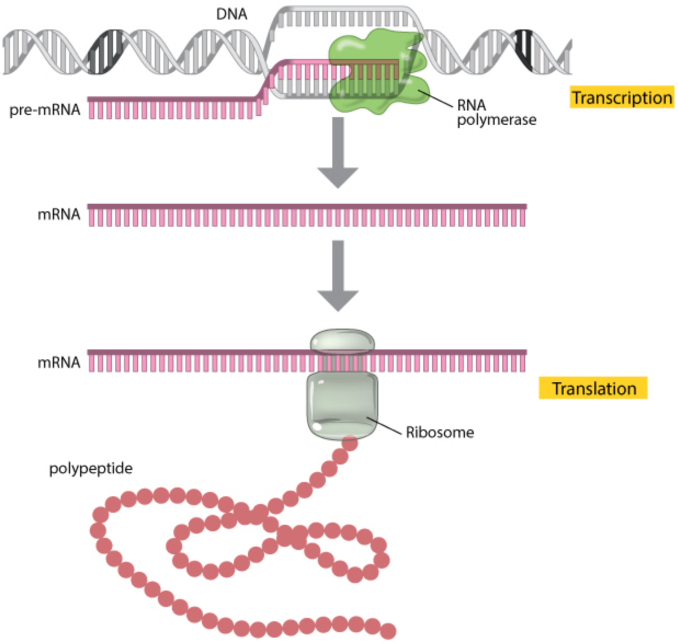

Figure 2.2: Transcription and Translation. Peptide chain represented traditionally as a necklace, amino acids being single beads. Source: https://www.nature.com/scitable/topicpage/translation-dna-to-mrna-to-protein-393

a peptide chain. Usually a peptide chain is at least around 250 amino acids long.

When amino acids are lined up to form a protein, they arrange themselves to form secondary structures. Two basic secondary structure types are known: alpha-helix, which is a coiled shape, and beta sheet, which is a zig-zag shape. Throughout the folding of the tertiary structure, amino acid residues link the distant pieces of a protein together.

I have very briefly introduced the process of protein folding but I did not mention why proteins are so important and what are they good for. Proteins are working components in our body and they have many functions in metabolism. For instance, they are the bricks that build bones and cells, they are transportation machines, DNA readers and interpreters, etc. All these various functions are due to the differences in proteins' shapes (structures). Protein haemoglobin, for example, has a specific shape that fits to the oxygen molecule, thus can function as oxygen transporter.

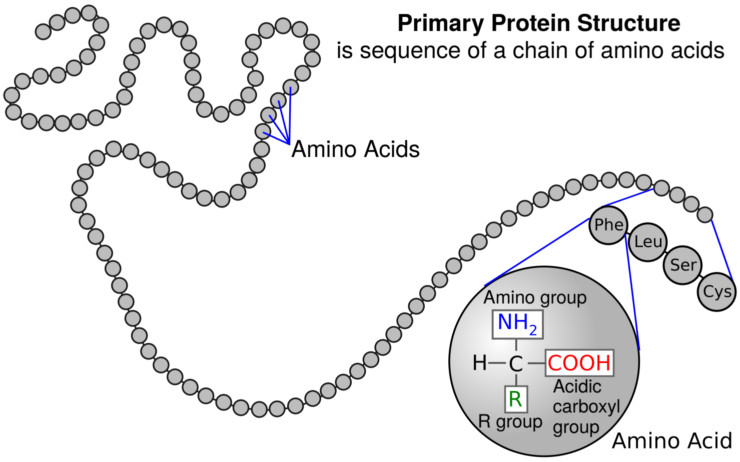

Figure 2.3: Peptide chain of amino acids, focus on one particular amino acid structure. Source: https://en.wikipedia.org/wiki/Amino_acid

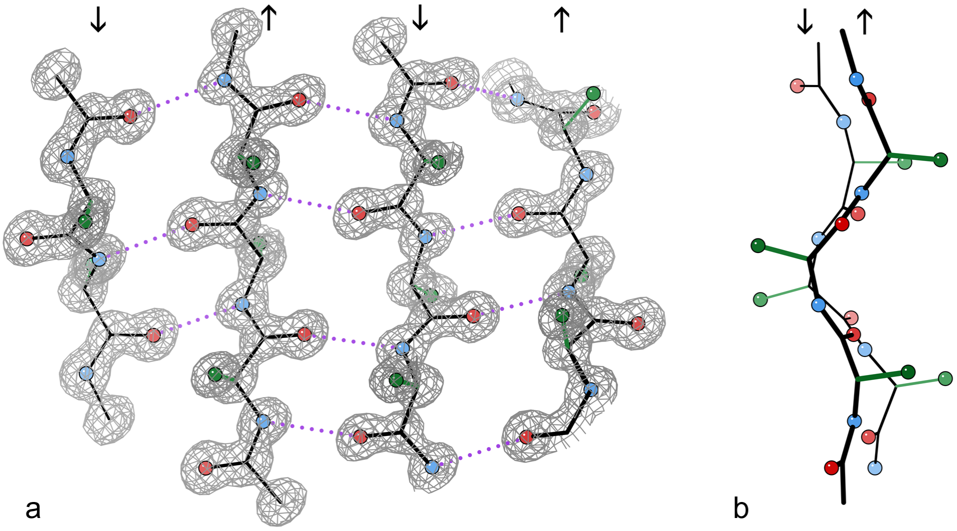

Figure 2.4: Example of a protein secondary structure: Beta-sheet fragment. a)front view, b) edge-on view. Source: https://en.wikipedia.org/wiki/Beta_sheet

The unsolved problem in current research in protein studies is the process of transition from the unfolded sequence into a unique three-dimensional conformation. As it was mentioned above, one single amino acid sequence may form different structures in different proteins. The following question unavoidably arises: why does a sequence fold in one way in one situation and in another in another situation? Or in other words: what is the relationship between sequence and structure? The question about the relationship between sequence and structure warrants research into thermodynamic explanations. A possible thermodynamic explanation would be that all possible constructions are tested to find the energetically most favourable one. But the amount of all possible construction of one peptide chain is so high that it remains a mystery to understand how evolution came up with stabilized protein structures. It was calculated by Cyrus Levinthal (Levinthal 1969) that a random search for protein structure would take up to 5 × $10^3$4 seconds, or 1.6 × $10^{27}$ years. The enormous difference between the calculated and actual folding times is called *Levinthal's paradox.*

The reader may ask at this point why are we even trying to understand the mysterious rules of protein folding? As was mentioned previously, protein studies might be very useful for pharmaceutic purposes. Many diseases are caused by folding errors. Understanding how proteins fold may help find compounds that will prevent errant folding. Understanding the process of protein folding may unlock the essential information researchers need to understand health conditions such as Alzheimer's disease or Parkinson disease, since it is known that the errant folding of proteins in the brain can lead to this conditions.

The current, mainstream research in protein folding is based on "homology modelling". That means, basically, that it is believed that the protein shape is completely determined by the sequence of amino acids (the peptide chain). Consequently, if we know a peptide sequence of an unknown 3D protein shape which is homological to a sequence which we already know the shape of, we use this shape as a model, a template for the 3D structure of the homologic sequence with an unknown shape. For the successful functioning of the aforementioned

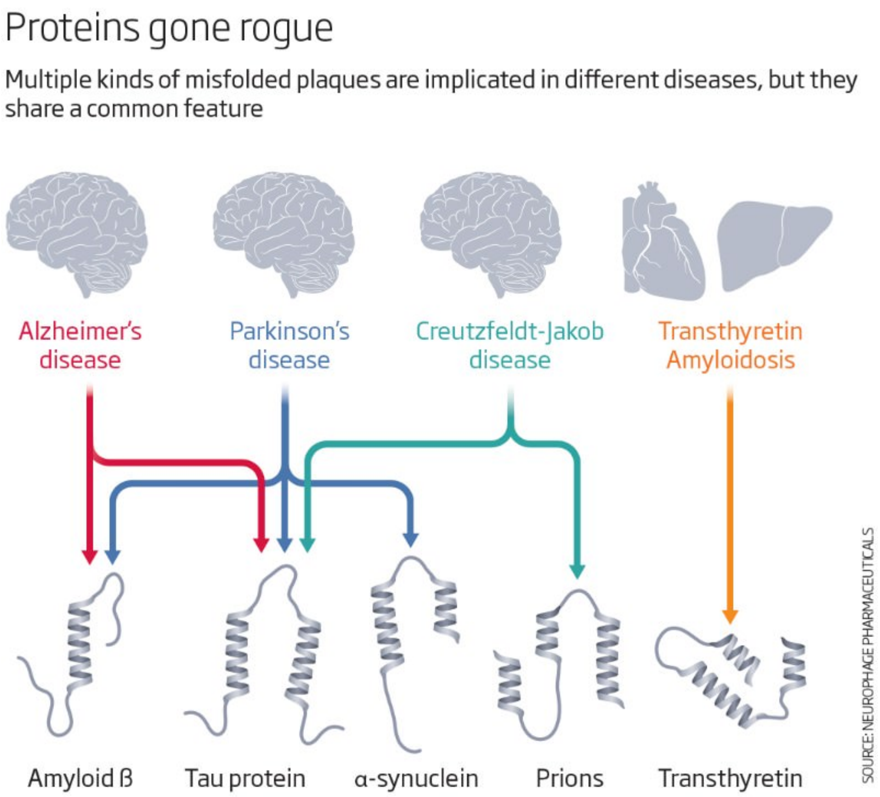

Figure 2.5: Diseases caused by errant folding. Source: https://www.newscientist.com/article/dn27921-universal-plaque-busting-drug-could-treat-various-brain-diseases

method, one has to start with the postulate that "the extent of the structural changes is directly related to the extent of the sequence changes" (Chothia and Lesk 1986).

Even though this method is widely used and has brought many results in protein's shape prediction it is not always applicable; many cases have been described which disclaim the one to one correspondence between the shape and the sequence. The prediction of a three-dimensional structure from sequence has proven to be extremely difficult. The possible reason why this method does not function is simply that not all amino acid residues in a sequence are responsible for the final fold, and thus it may be misleading to compare the percentage of sequence homology. "In fact, the percentage of dissimilar residues in proteins with similar structures may be surprisingly large – greater than 75%" (Kister 2015).

Another example of non-unequivocal correspondence between sequence and structure might be a case of the so-called solenoid proteins. These types of proteins were discovered as the simplest protein structures, the structural simplicity being due to the superhelically arranged repeated structural units (Kobe and Kajava 2000). The surprising fact about the class of solenoid proteins is that, repeating of basic structural units notwithstanding, singular solenoid proteins often consists of very dissimilar peptide sequences.

> Because solenoids are built of repeated structural units, one might expect these units to share sequence similarities. Indeed, many solenoids contain sequence repeats, although in other cases the sequence similarities are weak [e.g. protein farnesyltransferase, insulin-like growth factor-1 receptor] or nondetectable. (Kobe and Kajava 2000, p. 512)

I am not disclaiming the scientific merit of homology modelling, yet, given the proofs of non-one-to-one correspondence between sequence and structure, it seems improbable that this method alone could discover the rules for protein folding; a study of mere sequence is not promising. I mentioned the limits of homology modelling due to the fact that, for the sake of my argument, it is important to call attention upon the non-unequivocal relation between strings and proteins. This remark promotes my idea of a principle of structurality in the "protein language" which cannot be reduced to physical or chemical explanations in the protein folding process. Of course, chemical affinity and physical principles play an important role in the construction of a protein, yet what I want to stress is the fact that, besides chemical and physical factor, there is also an evolutionary process of "negotiation" of an arbitrary set of rules.

By the term "arbitrary" I mean the arbitrariness as it was defined by F. de Saussure (de Saussure 2011), in other words, by arbitrariness I do not mean an absolute randomness of possible protein structures, a randomness of set of rules which govern protein folding. The set of rules cannot be random, as far as it must be physically possible. That is; there isn't an unbounded set of possible options, in the same manner as in natural language, the arbitrariety of language rules is limited by acustic properties of sound chain and by anatomic properties of human articulatory system or human cognitive capacities etc. Thus, there is not an unlimited number of protein structures, quite on the contrary, the number of structures that occur in nature is rather close. A list of all known structures is evidenced in various databases, for instance a list of existing so-called “knot” protein structures is presented in the Protein Knot server (Kolesov et al. 2007).

As Kister remarked, the one and the same functional protein might be obtained from more than one amino acid chain. Likewise, the opposite case is possible: one and the same amino acid chain may perform more than one function and the function it will perform depends on the context, on the actual environmental surrounding: this is the case of the “moonlight proteins”. A phosphoglucose isomerase is one example of a typical moonlight protein:

> Take phosphoglucose isomerase (PGI), a glycolytic enzyme best known for its ability to convert glucose-6-phosphate into fructose-6-phosphate. The mammalian enzyme is now known to function as a neuroleukin (neurotrophic factor), an autocrine motility factor (AMF), a differentiation and maturation mediator for myeloid cells, and an implantation factor (in the ferret). There is now clear evidence that PGI is an important modulator of tumour progression and a target for cancer therapy. (Henderson and A. Martin 2011).

Cases like moonlight proteins can be useful for moderating strict reductionism in biology. By reductionism I mean scientific explanations based on the reduction to a lower level, for instance to explain chemistry by physical laws or to explain biology by chemical laws. The core idea of explanations in modern sciences is to explain natural phenomena from a reductionist point of view, reductionism being a scientific standpoint which is anchored in philosophical thinking since Descartes and is generally accepted by the scientific and also non-scientific community. Reductionism becomes controversial when applied to social and human sciences (think of the famous misuse of

Darwinism to promote racist ideas in the 19th and 20th centuries), nevertheless, reductionism might also become controversial in particular areas of the natural sciences, primarily in biology. Proteins are biological macromolecules composed of a folded peptide chain and characterised mostly by their function, which cannot be simply and exhaustively explained by the chemical properties of a peptide chain (reductionism); the fact that many peptide chains can acquire different functions depending on the actual context proves that the pure chemical explanations of protein function does not hold. These proteins are also called "geographical" or "moonlight proteins" (Henderson and A. Martin 2011; Huberts and van der Klei 2010; Jeffery 2005, 2014), "geographical" because of the topological dependency of the final function (e.g. a protein has one function when in a cytoplasm, another function when within a membrane, etc.) and "moonlight" as an analogy to a "moonlight job" or having more than one's day job. J. Dupré argues:

> The important property of proteins in biological cells is their ability to interact with other molecules. An important subclass, enzymes, catalyse chemical changes to such other molecules. But, first, being a particular kind of enzyme, a protease or a DNA polymerase, say, is a relational property among the enzyme, the substrate to which it bonds and the transformation that it catalyses. As it happens, a particular protein may have several kinds of enzymatic activity, and what it actually does depends on where it is in the cell and what else is in its vicinity ... This example, finally, exemplifies a central general claim I want to make against reductionist positions: reductionist methods explain how it is possible for an entity to have a particular capacity, but to understand what capacities it exercises, and even, I want to say, what capacities it actually rather than merely possibly has, require seeing the entity in a larger context. (**dupré2010**).

Dupré's point is that proteins in general, and "moonlight" proteins in particular, cannot be explained by naturalistic-reductionist perspective, if for no other reason than simply because of their very function in interaction with their surroundings. In other words, a protein as a three dimensional structure cannot be explained reductively by the chemical properties of the peptide chain (understood as one dimensional sequence) it is formed of, because it acquires its essence (function) only when the peptide chain is folded and only when it enters in a relationship with the surrounding molecules. Only the context determines and completes the final function—without the context, the protein fibre is ambiguous. The need for context becomes clearer by comparison with another linguistic analogy. Imagine a sentence without any linguistic or situational context, for instance the Italian sentence *La vecchia porta la sbarra* which have two meanings depending on the context[4]:

(1) La vecchia porta la sbarra
  a. [[La$_{\text{det}}$ vecchia$_{\text{n}}$]$_{\text{np}}$] [porta]$_{\text{v}}$ [[la$_{\text{det}}$ sbarra$_{\text{n}}$] $_{\text{np}}$]
  'The old lady brings the bar'
  b. [La$_{\text{det}}$ [[vecchia$_{\text{adj}}$porta$_{\text{n}}$]$_{\text{n}}$]$_{\text{np}}$ [ [la$_{\text{pron}}$ sbarra$_{\text{v}}$] $_{\text{vp}}$]
  'The old door bars her'

One can argue that my example is barely valid, since the sentence *La vecchia porta la sbarra* is an example of a very rare linguistic phenomenon, an amphiboly, wherein several polysemantic words happen to be placed together by accident, and consequently, my example is a language rarity rather than a general linguistic trait. But, as a matter of fact, polysemy is more present in the everyday use of language than one might think. Most lexical units (and in particular those with high frequency of usage) are polysemantic. For instance, the majority of highly frequented verbs are polysemantic. English verb "to carry" may acquire many different meanings depending on what words are in its proximity: carry about, carry along, carry away, carry forth, carry forward, carry back, carry in, carry on, carry off, carry over, carry up, etc. And the polysemantic nature of the verb is not exclusively related to phrasal verbs. According to Oxford English Dictionary (Stevenson 2010), there are even more than forty-three meanings of the verb "to

---

[4] This phenomenon was might be also considered as an example of the so-called bracketing paradox (Spencer 1988).

carry" which are not phrasal uses of the verb: to transport, to bear or take (a letter, message, report, news, and the like), to take by force, to cause to go or come, to extend or continue (a line, a piece of work) in the same direction to a specified distance, to win and many others. Similarly, in the Czech language, the verb *vést* (to carry) has more than thirteen different meanings (according to the dictionary *Slovník spisovného jazyka českého*). The polysemanticity of the Czech verb *vést* was commented on by P. Karlík (Karlík 1996). He argues that the polysemantic nature of a verb goes beyond the domain of semantics or lexicology. It is not only that the verb *vést* has many meanings in different contexts, but also the very syntactic or functional character of the verb changes. Karlík calls this a "syntactic verb". To understand better, let's have a closer look at examples presented by Karlík:

(2) Lendl vede
Lendl lead-INTRAN
'Lendl is winning'[5]

(3) Pavel vede Petra do školy
Paul lead-TRAN Petr-ACC to school
'Paul leads Peter to school'

(4) Vede se mi dobře
Lead-INTRAN REFL me-DAT well
'I'm doing well'

(5) Pavel vede pěvecký kroužek
Paul lead-TRAN sing-ACC class
'Paul leads a class of sing'

(6) Cesta vede do lesa
Path lead-INTRAN to forest
'The path leads to the forest'

(7) Poznání nás vede k pokoře
Knoweledge us-ACC lead to humiliation
'Knowledge leads us to humiliation'

---

[5] Lendl is a famous Czech tennis player.

(8) Minerálky nevedeme
Mineral lead-NEG-TRAN
'We do not sell mineral water.'

(9) Pavel vede rozmařilý život
Paul lead-TRAN capricious life
'Paul has a capricisous life.'

(10) Co to vedeš za řeči
What it lead-TRAN behind talking
'What are you talking about?'

The aforementioned examples of different uses of the verb "*vést*" are not purely lexical variations; in other words, the problem is not simply polysemy. As Karlík suggests, it is important to realize that one lexical item (*vést*) changes in syntactic structure in different linguistic contexts. The most evident manifestation of a change in the syntactic structure is the varying number of verb valence: whereas in the cases (2; 4) the verb is monovalental, in the cases (3; 7) the verb becomes trivalental and in cases (5; 9; 10) the verb is bivalental. The transitivity also changes in different contexts. Consequently, the diathesis relationship cannot act in the same way in all examples (3 - 10). Sentence (3) – *Pavel vede Petra do školy* – can be realized in both active and passive diathesis; the sentence:

(11) Petr je veden Pavlem do školy
Petr is led Paul-INSTR to school
'Petr is led by Paul to school'

is grammatical. On the contrary, the sentence:

(12) Lendl je veden
Lendl is lead-PASS
'Lendl is being won over'

is ungrammatical. A similar situation arises when we take into account the category of grammatical aspect: in the sentences

(13) a. Lendl vede
Lendl win-IMPER
'Lendl is winning'

b. Lendl vodí
   Lendl win-ITER
   ‘Lendl wins’

The second one is ungrammatical.

A similar situation can be found in the Italian language: the verb “*condurre*” (to carry) has at least six different contextual meanings (according to the Treccani dictionary). For example:

(14) a. condurre i figli a scuola
   ‘to lead children to school’
 b. condurre la moto
   ‘to drive a motocycle’
 c. condurre una linea
   ‘to draw a line’
 d. condursi bene
   ‘to behave well’
 e. condursi male
   ‘to behave badly’

I have illustrated briefly the concept of polysemy in three Indo-European languages and I tried to demonstrate that the polysemy as such is not a matter of language rarity; rather it is a very common language feature which may interfere with the syntactic structure of verbs or change the linguistic function of a word. Moonlight proteins, similarly to the Italian sentence *La vecchia porta la sbarra* are only extreme cases of the phenomena of polysemy, but the role of the context in language and in proteomics is present in the most common words and most common proteins.

# 3

# Thought experiment

## 3.1 Linearity and individuality

The remark I would like to make here is a semiotic one. So far, the way of understanding proteins has always been unidirectional, that means, from sequence to structure (sequence of amino acids determines the structure of the protein). In other words, the linear peptide chain has been comprehended as an entity formally and substantially determining the final protein structure. “Formally” because sequence order of singular amino acids is supposed to code for the final folds, to determine “positions of the chain where to fold”. “Substantially” because it is the peptide chain itself that the protein is made of. As a consequence, the peptide chain has been comprehended as an entity already articulated when entering in the relation with protein’s shape and function. To be precise, the peptide chain is usually treated as *linear*. But linearity is tricky, it is a term already involving very important feature of matter, that is, when I say that organic matter such as peptide chain is linear, what I am saying is that it is composed of individual units lined up one after another. To put it a different way, to be linear presupposes an existence of individual entities, distinct and separate individual features composing the entity that is to be linear.

What is surprising is that not only biologists (Monod 1972; Jacob 1970) but also semioticians (Deleuze and Guattari 1987) and linguists (Jakobson 1971a) treated the peptide chain as an entity pre-structured and articulated per se. But, I hold the opinion that there is another option of comprehending protein fibre. I propose a small thought experiment. Let's suppose the peptide chain as unarticulated mass, "a body without organs", an amorphous continuum. Even if it might seem unintuitive, the experiment is simple, one does not have to try hard be emancipated from the familiar image of a protein fibre as a chain of beads in a necklace. After all, a necklace is nothing but the habit of scientific schematization of biological entity which could be schematised in a different way, for instance as a simple line. Recall that there is divergence between scientific models and what they actually represent. Sometimes it is hard to differentiate between scientific models and reality, yet one should keep in mind that scientific models are matter of vogue in the contemporary scientific paradigm. I did not use the term "vogue" accidentally; sometimes scientific models resemble real artistic masterpieces and correspond with the scientist's aesthetic taste (e.g. Haeckel's famous biological drawings).

But, let's go back to our experiment. To imagine a peptide chain in a non-linear way, that means, without particular individual amino acids composing it, the only step one has to do is to comprehend peptide chain in a molar (and not molecular) way. This model, the "molar peptide model", will allow us to be liberated from the sequentiality of the peptide chain as was pre-established.

The sequentiality of the amino acid chain is its characteristic feature and there is a correspondence between amino acid sequences and particular protein folds, as demonstrated in (Chothia and Lesk 1986). But, what is important to realize is that this sequentiality probably does not determine structure in the sense of a generating machine. It is only a means to express and "code for" structures, but this means it cannot be

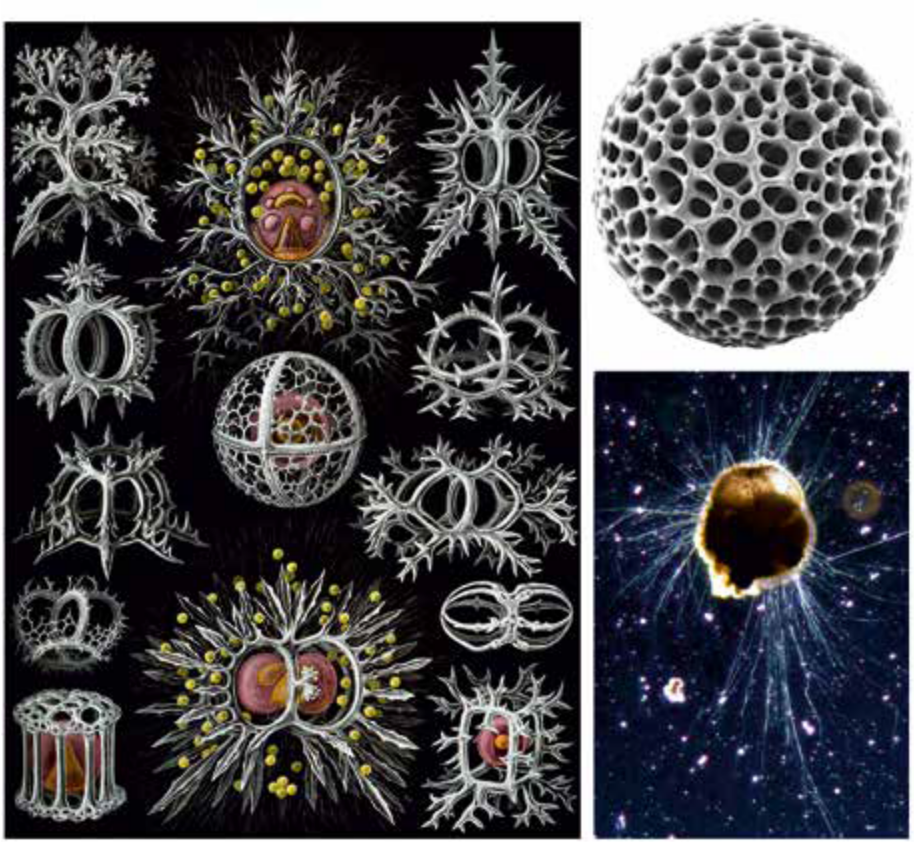

Figure 3.1: Diversity in representation of a cell: Haeckel´s drawings from 1862 and a real cell (Markoš and Cvrčková 2018).

perfect because of its very nature of linearity. Analogically to natural language, word order “codes for some meanings” within a sentence, as in, for instance, syntactic categories or thematic relations (e.g. in English, the first position in a sentence codes for subject and the first position after verb codes for object, etc.), but word order does not have the power to determine or generate the meaning, it is only one of possible means of coding for this meaning (and, evidently, not perfectly reliable, since it is not universal for all languages and for all pragmatic situations: consider the theory of functional sentence perspective). As is commonly known, languages have many different ways of expressing this kind of meaning, for example a flexion and case system (Czech) or affixes (Papuan languages). In the Papuan language Yimas, the thematic relations are codified by double simultaneous use of affixes on substantives and verbs (Foley 1986, p. 94). In the sentence

(15) nama-t ura-ŋk nar-maŋ kɨ-n-ŋar-umpum
man.I:PL coconut.VI:SG woman.II:SG VI:SG-II:SG-give-PERF-I:PL
‘A woman gave a coconut to men’

we observe that the suffix of each substantive is repeated as a verbal prefix. The order of verbal affixes then determines the syntactic categories or thematic relations within the sentence. The most external prefix indicates a direct object (i.e., coconut), a prefix closest to the lexical morpheme (respectively the root of the word) indicates the agent (woman). The postradical verbal affix (if present) refers to an indirect object or the beneficiary.

To summarize, in light of all these language facts, we can say that sequencing is meaningful. It is a set of rules that codes for structure, yet it can be misleading to exchange the syntactic rules of a language for language itself. A language or a code, semiotically speaking, can only be comprehended as an inseparable connection between expression and content, syntactic rules being a kind of “s-code” are of no use to semiotics if not studied in connection with meanings it codes for. A code, in general, can be defined as a correspondence between two independent worlds. Umberto Eco distinguished two basic types of codes: “s-codes” and “codes”, wherein “s-codes” only represent internal set of rules without any correspondence to another world (another “s-code”). By way of example, mathematics, phonological code, and moral code are “s-codes”. “Code”, on the other hand, is a real code that represents a set of rules that create correspondence between two “s-codes”. For example, the natural language is a “code”, wherein one “s-code” (phonological code) corresponds to another “s-code” (semantics). My understanding of syntax goes in the direction of C. Morris, who defined syntax (syntactics) as the study of relations among the members of a single set (Morris 1971, pp. 28–31).

What is important to note is that there is no natural relationship between the two worlds (phonological and semantic), the relationship is purely conventional and this is why one has to learn a language to understand it, as well as the rules which keep the meanings together with the sounds. Consequently, there is no way to understand a language intuitively. This particular feature of language was called arbitrariness by F. de Saussure (de Saussure 2011).

It can be said that the relationship between the protein fibre and the function of a protein is also a matter of arbitrariness, in other words, a result of evolutional convention and not a result of physical necessity.

Of course, arbitrariness by itself is not a constitutive principle of a protein structure. Quite on the contrary, physical laws (physical laws being a counter-example of arbitrariness) of thermal stability determine the structure of folded peptide chain. But still, there are features of protein folding which are related to the principle of arbitrariness, starting with the choice of the twenty amino acids and ending with Levinthal's paradox.

In order to better anchor the understanding of the peptide chain as an "s-code", I turn to J. Monod:

> With the globular protein we already have, at the molecular level, a veritable machine- a machine in its functional properties, but not, we now see in its fundamental structure, where nothing but the play of blind combinations can be discerned. Randomness caught on the wing, preserved, reproduced by the machinery of invariance and thus converted into order, rule, and necessity. A totally blind process can by definition lead to anything: it can even lead to vision itself. In the ontogenesis of a functional protein are reflected the origin and descent of the whole biosphere. And the ultimate source of the project that living beings represent, pursue and accomplish is revealed in this message- in this neat, exact but essentially indecipherable text that primary structures constitute. Indecipherable, since before expressing the physiologically necessary function, which it performs spontaneously basic make-up it discloses nothing other than the pure randomness of its origin. (Monod 1972, p. 98)

What Monod described in this part of his famous book is in fact nothing more than the difference between "s-codes" and "codes", and the semiotic need to comprehend expression as inherent to content (meaning). In other words, for Monod the "indecipherable text" of protein fibre represents *expression*, the "necessary function" represents *content* and "randomness" is nothing but another term for *arbitrariness*. I consider the aforementioned quotation by Monod to be essential, since it is one of the rare examples of semiotic thinking in biology: the inseparability of expression from content (and undecipherability of expression by itself) as one of the basic characteristics of protein code.

Now let's proceed with our thought experiment. Let's suppose that the sequentiality of amino acids within a peptide chain does not generate the protein structure, but, conversely, it exists only thanks to and because of the protein structure: "indecipherable before expressing the function" means that before the birth of a functional protein the expression itself has nothing to code for and thus, we cannot even consider it as having semiotic existence. That means that the individual existence of single amino acids emerge only thanks to protein structure: thanks to the folds that give birth to the structure. Until that moment the peptide chain is, let's say, an amorphous continuum. In order to understand deeper this idea, it is necessary to introduce Peircian term of continuum.

## 3.2 The Folding of a Continuum

The contrast between continuity and discontinuity is present in every semiotic theory. Whether we consider Hjelmslevian "purport" or Saussure's "amorphous mass", the passage from continuity to discontinuity constitutes the underlying concept at the very heart of semiotic theory, creating discrete and meaningful units that enable us to conceive the world, to orient ourselves inside this world and finally to communicate about things in this world. The passage from unformed unanalysed sound (expression) and unformed unanalysed thought (content) to articulated and formed substances has been the concern of semiotic inquiries since its beginnings. The discontinuity induces the emergence of signs and meanings.

Despite elegant European structuralist theories, Peirce's notion of continuum is more complex than that. A continuum does not disappear with arrival of articulated signs; on the contrary, a continuum (Synechism) is only created by signs. In Peirce, a continuum is rather a term for connectedness more than a term for lack of form. But it is more than that. A continuum for Peirce has to do with Peircian logic, in particular, with the possibility of violating the law of excluded third.

> Now if we are to accept the common sense idea of continuity (after correcting its vagueness and fixing it to mean something) we must either say that a continuous line contains no points or we must say that the principle of excluded middle does not hold of these points. The principle of excluded middle only applies to an individual (for it is not true that „Any man is wise" nor that "Any man is not wise". But places, being mere possibles without actual existence, are not individuals. Hence a point or indivisible place really does not exist unless there actually be something there to mark it, which, if there is, interrupts the continuity. (Peirce CP, § 6.168)

Actual existence does not occur unless there is something that interrupts the continuity, which is why, within the continuity, the principle of excluded third is not violated: if actual existence does not exist, what only exist within the continuity is the potentiality ("may be", the conditional replaces imperative) and potentiality means the coexistence of contradictory facts. Peirce explains his theory of continuum also in a mathematic way, as a dispute of the famous Dedekind cut:

> Personally, I agree entirely with James, against Dedekind's view; and hold that there would be no actually existent points in an existent continuum, and that if a point were placed in a continuum it would constitute a breach of the continuity. Of course, there is a possible, or potential, point-place wherever a point might be placed; but that which only may be is necessarily thereby indefinite, and as such, and in so far, and in those

> respects, as it is such, it is not subject to the principle of contradiction, just as the negation of a may-be, which is of course a must-be, (I mean that if "S may be P" is untrue, then "S must be non-P" is true), in those respects in which it is such, is not subject to the principle of excluded middle. (Peirce CP, § 6.182)

If we imagine a continuum as a continuous line, Dedekind's comprehension of cutting a line consists in that the cutting point, the border between parts in relations to each other (two parts of a line), must always be assigned to only one of the two parts, so as to be able to be always reduced to a determined individual entity. For Peirce, the cutting point does not belong to either of the two parts, or rather, belongs to both of them (Paolucci 2004, p. 125).

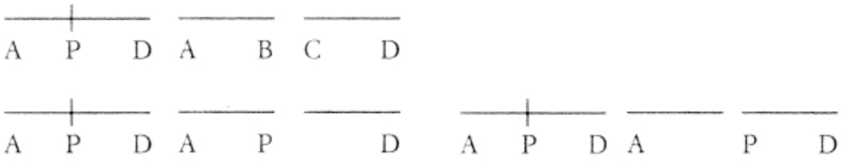

Figure 3.2: Cutting a line by Peirce and by Dedekind: the Peircian cut is represented by the upper line (AB, CD) and Dedekind's cut is represented by the lower line (AP, D or A, PD). (Paolucci 2004, p. 125).

If we imagine the continuum as a continuous surface, the individual points belonging to the border line (cutting line) must always be, according to Dedekind, assigned to only one of the two parts in the relationship (A or B). On the contrary, for Peirce, the division of a continuous surface into two parts leads not only to two distinguished parts, but also to a third part considered to belong to both parts, in other words, the third part is undecided and exists in the conditional mode, as possibly to be either A and either B.

As was illustrated above, the notions of continuum in European structuralism and in Peircian semiotics differ a lot. In structural theories, binary relations play crucial role and it is binary relations which constitute the discontinuity of amorphous masses. In Peirce, on the contrary, binary relations are not considered to be part of his semiotic theory, only triadic relations give birth to semiosis and the whole sign process. In structural theories, the victory of dis-continuum over continuum is celebrated. In Peirce, the discontinuity in the aforementioned sense does not exist, yet, in Peircian theory binary relations also have their place. Binary relations (discontinuity) are for Peirce a kind of broken continuity, but the important difference between Peirce and Saussure or Hjelmslev is that for Peirce, the broken continuum conserves, notwithstanding being broken down, its continuity. How this breaking down of a continuum without losing the continuity is possible was elegantly explained by Paolucci (Paolucci 2004). Paolucci proposed to distinguish between the terms of cut and fold. In fact, what Dedekind's model of a cut represents is simply a cut. Contrarily to this, Peirce's model of "cut" should rather be replaced by the notion of *fold*. Because only the folding of a line or of a continuous surface breaks down the continuum and at the same time conserves the continuity. Furthermore, the points of the folding line (contrarily to cutting line) are undecided, they exist in the mode of the conditional, of may be. In fact, they are not points in the sense that they are not single, individually existing entities. G. Deleuze was also fascinated by the notion of the fold, to which he dedicated the essay *The Fold. Leibniz and the Baroque* (Deleuze 1988). Deleuze retains that the fold, not the point, is in fact the smallest element of a matter, his observation being similar to Peirce's critique of Dedekind's cut:

> The unit of matter, the smallest element of the labyrinth, is the fold, not the point which is never a part, but a simple extremity of the line. That is why parts of matter are masses or aggregates, as a correlative to elastic compressive force. Unfolding is thus not the contrary of folding, but follows the fold up to the following fold. (Deleuze 1988, p. 6)

Deleuze compared the process of the folding of the continuum to origami, the art of paper folding. Folding, unlike cutting, preserves the continuity of the paper sheet. Take an origami crane, unfold it carefully and flatten it out; it will become an unarticulated square sheet of paper. Take a protein and put it in an organic solvent, or heat it; it will dissolve the bonds responsible for folding, it will become a line again. When a protein is denatured, secondary and tertiary structures are altered, yet the bonds of the peptide chain, the bonds between amino acids, remain intact. Proteins can return to their folded (or so-called "native") state again. This process is called renaturation, which is similar to an unfolded sheet of paper which one can fold again to regain the initial paper crane.

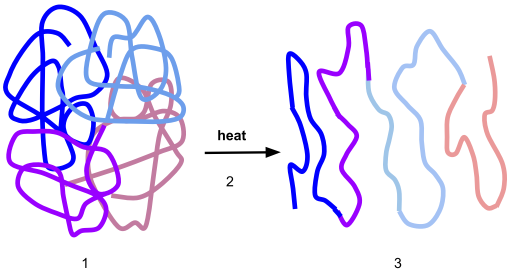

Figure 3.3: Denaturation of protein. Source: https://en.wikipedia.org/wiki/Denaturation_(biochemistry)

Protein folding is a perfect example of a folded continuum which, even when creating binary relations (between one and another part of a folding point), preserves its continuity since it does not break down peptide bonds between amino acids. A classic example of denatured protein is represented by boiling eggs. Egg whites contain the protein albumin. Fresh egg whites are transparent and liquid. Cooking egg whites turns them opaque; this is the result of denaturised albumin fibres that interconnect between themselves, thus creating a solid opaque mass.

I have introduced the basic difference of understanding continuum and continuity within the frameworks of Peirce and in structuralism. The two theories approach the notion of continuity differently; nonetheless, they are not incompatible. Paolucci explains how the coexistence of continuity and discontinuity is possible, and how structural binary oppositions are possible even within the Peircian theory of continuum. The idea of folding a continuum is that when the continuum is folded, it never breaks its continuity, it is just folded.

> From continuum and in continuum, without ever interrupting continuity, it is possible to create discontinuity

> effects that give rise to opposition relations.[6] (Paolucci 2004, p. 135).

Yet these binary relations of opposition are identifiable only when the continuity is folded, that is, they are a product of the fold of the continuity and not vice versa, the continuum is not generated by binding together binary relations. In fact, as Peirce demonstrated, we can only get dyads and monads from a triad. How can these be understood in terms of protein folding?

A single peptide chain represents an inconceivably enormous amount of possibilities of potential structures (remember Levinthal's paradox: at least $5 \times 10^{47}$ possibilities and probably many more). When one sees a peptide chain, they see nothing other an "infinity" of possible structures. Thus a chain is nothing more than mere possibility (firstness) without concrete identity (structure) and is potentially constitutive of an almost infinite amount of possible structures (Levinthal's paradox). For this reason, it cannot be said that sequence determines structure, since sequence determines a potentially infinite number of structures. We should say, on the contrary, that structure determines sequence, as far as only when the structure is given can the sequence adopt its identity. The amino acid residues of the final fold are decided as binding places (relates), thus, we can say that amino acids adopt their identity. Compare this with the following passage by Peirce:

> When we say that of all possible throws of a pair of dice one thirty-sixth part will show sixes, the collection of possible throws which have not been made is a collection of which the individual units have no distinct identity. It is impossible so to designate a single one of those possible throws that have not been thrown that the designation shall be applicable to only one definite possible throw; and this impossibility does not spring from any incapacity of ours, but from the fact that in their own nature those

---

[6] In the original: "Dal continuum e nel continuum, senza mai interrompere la continuitá, é possibile creare degli effetti di discontinuitá che danno origine a relazioni di opposizione."

> throws are not individually distinct. The possible is necessarily general; and no amount of general specification can reduce a general class of possibilities to an individual case. It is only actuality, the force of existence, which bursts the fluidity of the general and produces a discrete unit. (Peirce CP, § 4.172)

Only actuality produces discrete units, and only protein structure produces amino acids as discrete units of a particular fold. Only when "the die is cast" the throw acquires its particular existence, identity, and until that moment it exists only as a possibility. So to say, a peptide chain exists only as a possibility, and its particular amino acids as well. Unless they are folded, they do not exist as singularities, which is why proceeding from sequence to structure does not give desirable results in predicting protein structures.

A peptide chain is like a continuum, constituted by strong chemical bonds which never dissolve and a peptide chain folds, potentially unfolding and refolding again. Moreover, only when continuum is folded is dis-continuity generated; only at this point we can identify amino acids as "single points" responsible for the folds. Paolucci's lesson on folding a continuum brings about several important points.

Firstly, the notion of the *folding* (and not cutting) of a continuum guarantees the possibility of transition and coexistence between Peircian interpretative semiotics and structural binary semiotics. Binary relations or dyadicities are possible within the theory of a continuum composed of triadic relations. Even so—and this is a very important point—dyadic relations are only possible when generated from a triadic relation, only by the folding of a triadicity is a dyadicity possible. That means, thanks to the fold, even when binary relations are created, this creation never breaks down the continuity of the triadicity.

> From continuum and in continuum, without ever interrupting continuity, it is possible to create discontinuity effects that give rise to opposition relations. (Paolucci 2004, p. 135)

Secondly, direction plays crucial role in passing from continuity (triadicity) to discontinuity (dyadicity). As Peirce accented in several places, a triad can generate dyads or monads but a transition of opposite direction is never possible: a triad cannot be generated from dyads since the triad is the primitive relative indecomposable to smaller units of relation, this principle is also called Non Reduction Theorem by Peirce (I will return to this theorem in the next chapter). Likewise, Saussure, in his posthumously published writings, wrote that the linguistic form is obtained from the idea (de Saussure et al. 2002, p. 34).

## 3.3 From triadicity do dyadicity

Now, I would like to continue with my thought experiment and to proceed to the application of the theory of a folded continuum to the "organic strata". As was previously proposed, a peptide chain could be, for the sake of the semiotic interpretation of biological phenomena, comprehended as continuous line, unarticulated mass, or undecided possibility. The peptide chain is, in our experiment, a "body without organs". It is a "body without organs" unless folded. Once it folds, the continuum is broken by the action of folding and dyadicity emerges; folds are points of frontier. Possibility turns into necessity: a protein has its shape and it cannot have had another shape, at least unless it is denatured and refolded again. Consequently, we can talk about the dyadic relations of expression by which content is encoded. We can talk about amino acids as discrete units of organic expression which code for the organic content represented by a protein's function, or by the protein's shape, because, let's say, a protein structure and its function are coextensive notions.

Now, what was this thought experiment good for? Well, only if we succeed in liberating ourselves from a linear understanding of the peptide chain can we confront the relation between content and expression. In no semiotic theory does expression generate content. Content and expression are two separate planes brought together thanks to, let's say, habit or convention. Of course, we can estimate what the meaning of a given expression is, but this is not the same as meaning generating.

Additionally, we can estimate what the meaning of a given expression is only if we know the expression rules, that is, only if we know the grammar. In other words, only if we know in what manner within a given code expression units relate to content units can we estimate what the actual meaning is.

Take this scenario into consideration: I receive a message from my friend saying, “A big pink unicorn will come for a dinner tomorrow”. The sentence, “A big pink unicorn will come for a dinner tomorrow”, is an expression I never heard before, yet I can estimate what this expression means. I can estimate its content because I know the rules of the code, I know the grammar of English language, therefore I know that the “big pink unicorn” is a subject and I know that “tomorrow” is a kind of shifter, so I am able to localise the meaning in the time and I can understand that tomorrow someone will come for a dinner. I can suppose the sentence has probably a metaphorical meaning, or I suppose it is a joke. But, if I hear a sentence in a Papuan language, even if I have a dictionary so that I can translate every single word of the sentence, I simply cannot catch the meaning of it. Thus, by analogy, since we do not know the mystery of the protein folding code, we cannot understand the very connection between the amino acid sequence and the final shape. Of course, the analogy is to be understood in terms of an isomorphism of relations, that is, the relation between sound (a phonetic chain) and meaning is isomorphic to the relation between amino acid stings and a protein’s shape. This fact, however, does not imply that linguistic meanings are comparable to protein shapes or protein functions. The notion of meaning in biology is very problematic and I will return to it in the next chapter. For the moment, and for my purposes, it is not important to give a precise definition of meaning in biology, only the relation between amino acids and a protein’s shape is important for my argument. The question I would like to pay attention to is how to decipher a code, or how to construe a grammar of an unknown language. To construe a grammar, knowledge, no matter how detailed and complex, of only one level (expression level), will never be sufficient. This is the reason why Chomsky, in the 1960s, introduced the concept of Deep Structure into his theory, which was supposed to be a kind of semantic level of language or the meaning of the sentence[7]. And, this could probably also be the reason why protein shape prediction by sequence homology and similar methods, no matter how elaborate and how complex, always has faults.

Thus far, we have seen that content in a grammatical description is inseparable from expression. The reason is simple; this is because we have expressions to express contents. If not, they would have no reason to exist. Consequently, to understand the rules of expression we have good reasons to start with understanding content.

This calls to mind Searle's famous argument about the Chinese room. One point of Searle's Chinese room argument is that pure syntax will never lead to an understanding of semantics[8]. This idea was attacked by Rapaport (Rapaport 2007), who argued using Helen Keller's true story. Helen Keller was born with the ability to hear and see, yet due to an illness at 19 months she became deaf and blind. Despite her handicap, she succeeded to learn language and to speak, thanks to help of her teacher, Miss Sullivan. She was the first deaf and blind person to earn a bachelor's degree. She learned to speak by way of tactile spelling on the palm of her hand. According to Rapaport, "growing up deaf and blind locked her in a sort of Chinese Room" (Rapaport 2007, p. 2). Nonetheless, being locked in the "Chinese room", she learned how to communicate with the outside world and this, for Rapaport, was a living example of the possibility to escape from a Chinese room, that is, to comprehend semantics by no other means than a pure syntax. But Rapaport's argument does not work, since Hellen Keller was not actually locked in a Chinese room. She grew up in the same way as

---

[7] My reference to the generative syntax here is just a short remark and I apologise for the superficiality of this reference. In fact, it is not completely coherent with the current theory to equal the notion of Deep Structure with semantic interpretation. Starting from 1970s, Surface Structure also determined, partially, semantic interpretation, because phenomena like topic and focus impact on semantics yet are transformational (Chomsky 1970). My point was, in any case, just to pay attention to the fact that also Chomsky's syntactic theory, in the end, cannot avoid to deal with semantics (semantic interpretation).

[8] Of course, a definition of the term "semantics" would be appropriate at this point. I will focus on this topic in the last part of the thesis. For the sake of clarity, I am using the term "semantics" at this place in the sense as it is used by (Rapaport 2007) in his argumentation.

a non-handicapped baby, in connection with the outside world where she could touch things and she could smell things. She was not in a Chinese room; she had only a limited way to perceive the world, or, rather, a different way to perceive the world. But, the mechanism by which she understood language was the same as normal kids: by associating expressions with meanings. It is not true that she only knew the pure syntax (the spelling of characters on her hand), she also knew the things the expression referred to since she had touched them, smelled them and she had felt the vibrations. The famous moment of realizing that entities from a non-linguistic reality are connected to the spelled words testifies to the fact that Keller only recognised and apprehended linguistic rules when she succeed in connecting the contents to the expressions: she understood the word "water" the moment she was touching it.

> Someone was drawing water and my teacher placed my hand under the spout. As the cool stream gushed over one hand she spelled into the other the word water, first slowly, then rapidly. I stood still, my whole attention fixed upon the motions of her fingers. Suddenly I felt a misty consciousness as of something forgotten, a thrill of returning thought; and somehow the mystery of language was revealed to me. I knew then that "w-a-t-e-r" meant the wonderful cool something that was flowing over my hand. (Keller 1903, p. 67)

Rapaport's example of Helen Keller is in fact a counter-example of what he was trying to prove. Keller did not come to semantics by pure syntax; she did not understand the syntax per se before understanding the meanings it referred to. Until the moment recalled in the water story, she was not able to comprehend very syntactic rules. Rapaport was wrong, she had not mastered the syntax, and she only repeated the spelling without understanding it. She memorised some of the spellings, yet memorising does not equal understanding.

> I did not know that I was spelling a word or even that words existed; I was simply making my fingers go in monkey-like imitation. In the days that followed I

> learned to spell in this uncomprehending way a great many words. (Keller 1903, p. 35)

Keller did not come to semantics from pure syntax, quite on the contrary, she came to syntax from semantics: she experienced the meaning by touching it and only after she conceived what the nature of spelling was and from that moment she learned very quickly all other words. The message of our thought experiment with peptide chain is the same as for case with Helen Keller: put simply, in a grammatical description, the direction matters. Firstly, one has to know the meaning and only after one can decipher the syntactic rules of an unknown language. First, the continuum folds, afterwards the dyadicities emerge. Dyadic relations do not generate triadic relations and sequentiality (linearity) does not generate dimensionality simply by folding itself. On the contrary, the continuity of a line (triadicity) generates sequentiality (dyadicity) by folding itself. The trick is to see the difference between a continuous line and linear (unarticulated) line (already articulated).

If we equate the sequence of amino acids as being expression and protein's shape (or function) as being content, we should start our grammatical description from the content. Because, from a semiotic-structural point of view, not all amino acids in a protein fibre that interest us, but only those amino acids which are folding points responsible for the protein's shape. All the rest are only the material, and material is not the grammar.

We can draw a phonological analogy for the sake of better understanding. When distinguishing between phonemes B and P, between voiced and voiceless consonant, the only thing that matters is the folding point of the continuity of sound. When I need to distinguish between Italian "palla" (ball) and "balla" (to dance, 3rd person sg of present indicative), the only distinctive point that draws folding line is important, all the rest is the pure phonetic material that does not play role in distinguishing the meaning.

But, how do I know where the pure phonetic material starts and where it ends? I only can know this once I know the meaning of Italian words

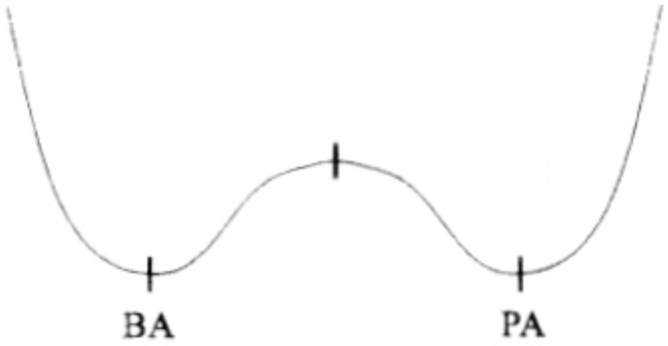

Figure 3.4: Folded phonological continuum (Paolucci 2004, p. 136).

"palla" and "balla". Phonemes make sense only because they differentiate meanings. Let's look at a better example. Imagine a situation of a Slovak speaker who does not distinguish between open and close vowel O. Slovak language does not have open and close vowels, so the Slovak speaker only hears the same word when an Italian speaker pronounce b/ɔ/tte (blows) and b/o/tte (barrel). Now the Slovak speaker starts to frequent a course of Italian language and they are taught that in Italian language they have this kind of a phonological minimal pair: /ɔ/ vs. /o/.

From this moment on, the Slovak speaker is able to hear the difference between /ɔ/ and /o/. But, before understanding the various meanings which the minimal pairs may distinguish (barrels and blows), they could not hear the difference.

Now imagine that the Italian teacher is trying to explain in his class the difference between b/ɔ/tte and b/o/tte (barrel)and he says: "Botte" with open O means "sud" in Slovak. The Slovak student will ask: /sud/ (barrel) or /su:d/ (court of justice)? And the Italian teacher will ask what the difference is, since Italian phonologic system does not differentiates between short and long vowels. Only a difference in content leads us to comprehend the difference in expression, that is, to comprehend the form or the rules of expression.

Now I can go back to folding a continuum, as linguistic expression—or phonology level—is generally understood as a plane composed of dyadic relations. Phonology is perfectly describable by binary opposition (open-close vowel, voiceless-voiced, etc.) But, it is describable by dyadic relations only because it was previously generated from meaning (content plane). It was the content, the differences of meaning, which lead to generating the dyadic relations of an expression plane. It was the triadic content that potentiated one to hear the dyadicity of expression. Until connected with content level, expression level was not dyadic, it was simply continuous, meaning also triadic.

To summarize, we have the continuum of an amino acid (purport or triadicity) chain, we have the linearity of an amino acid chain (substance or dyadicity) and we have the connection with content (function of protein) as a folding point which makes the transition from the triadicity of peptide chain to the dyadicity of an articulated (folded) peptide chain. I admit that the direction from triadicity to dyadicity is very unintuitive, in the context of proteins especially. It is especially unintuitive in the context of proteins because of the biological notion of protein synthesis starting with the genetic code and going from nucleic acids through amino acids to proteins. In other words, it goes from one type of linearity through another type of linearity to the final dimensionality of a protein. This optic of unidirectionality is applied not only in biology but also in semiotics and biosemiotics. But, to understand better, let's have a closer look to the classical understanding of the genetic code and protein synthesis.

## 4

# Code and Encyclopedia

## 4.1 Genetic code and semiotics

In biology, the term “genetic code” is understood as a table of 64 codon triplets of nucleic acids specifying amino acids or “STOP” marks. The below table is only a schematisation of real connections between nucleic acids and amino acids which exist in nature in the form of strings (or folded strings).

Table 4.1: Genetic code by Bolshoy (Bolshoy et al. 2010, p. 11).

| Amino Acid | mRNA codons |
|---|---|
| Alanine (Ala) | GCA, GCC, GCG, GCU |
| Arginine (Arg) | AGA, AGG, CGA, CGC, CGG, CGU |
| Asparagine (Asn) | AAC, AAU |
| Aspartic acid (Asp) | GAC, GAU |
| Cysteine (Cys) | UGC, UGU |

| Amino Acid | mRNA codons |
|---|---|
| Glutamic acid (Glu) | GAA,GAG |
| Glutamine (Gln) | CAA,CAG |
| Glycine (Gly) | GGA, GGC, GGG, GGU |
| Histidine (His) | CAC, CAU |
| Isoleucine (Ile) | AUA, AUC, AUU |
| Leucine (Leu) | CUA, CUC, CUG, CUU, UUA, UUG |
| Lysine (Lis) | AAA, AAG |
| Methionine (Met) | AUG [9] |
| Phenylalanine (Phe) | UUC, UUU |
| Proline (Pro) | CCA,CCC,CCG, CCU |
| Serine (Ser) | AGC, AGU, UCA, U CC, UCG, UCU |
| Threonine (Thr) | ACA, ACC, ACG, ACU |
| Tryptophan (Trp) | UGG |
| Tyrosine (Tyr) | UAC, UAU |
| Valine (Val) | GUA, GUC, GUG, GUU |
| Stop codons | UAA, UAG, UGA |

In biosemiotics, how the “genetic code” not only acts as a correspondence between two organic levels, but also a proper semiotic understanding of the process of semiosis is the point of interest. A semiotic explanation of life processes and life itself is the main task of biosemiotics. A standardized semiotic definition of signs and their meanings in the genetic code has not been established so far, so we can only talk about various views on the genetic code without any coherent unification in the approaches or terminology (Barbieri 2003, 2007, 2011; Emmeche 2011; Jakobson 1971a; Searls 2002; Stegmann 2015; Trifonov 1988). But, in general, the semiotic definitions of the genetic code do not differ a lot from those biological: there are two set of units brought together by the code: a DNA (or RNA) string as a set of nucleic acids on one hand and a peptide chain as a set of amino acids on the other hand. That means the understanding of the code is rela-

---

[9] At the beginning of a gene, AUG has the function of the start codon.

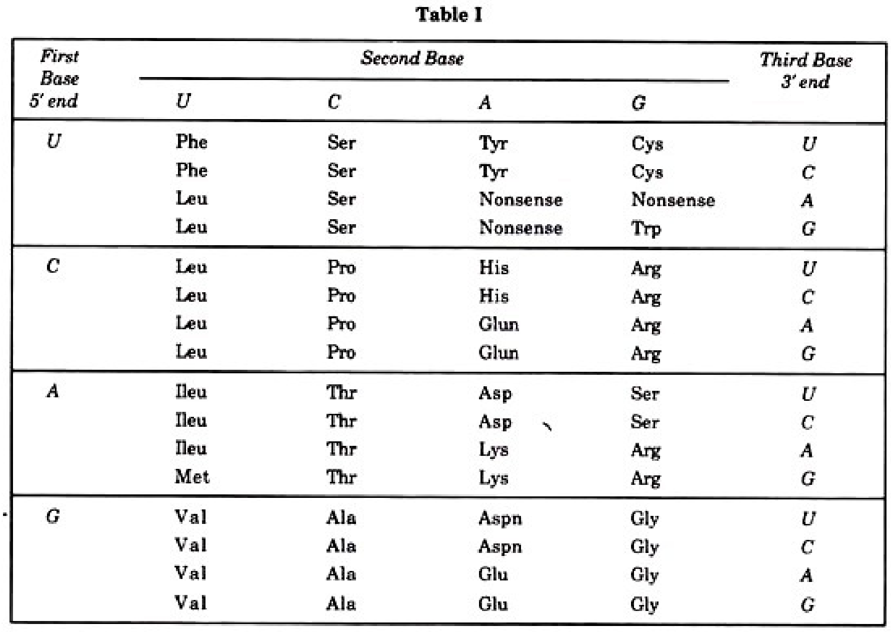

**Table I**

| *First Base 5' end* | *Second Base* *U* | *C* | *A* | *G* | *Third Base 3' end* |
|---|---|---|---|---|---|
| *U* | Phe | Ser | Tyr | Cys | *U* |
| | Phe | Ser | Tyr | Cys | *C* |
| | Leu | Ser | Nonsense | Nonsense | *A* |
| | Leu | Ser | Nonsense | Trp | *G* |
| *C* | Leu | Pro | His | Arg | *U* |
| | Leu | Pro | His | Arg | *C* |
| | Leu | Pro | Glun | Arg | *A* |
| | Leu | Pro | Glun | Arg | *G* |
| *A* | Ileu | Thr | Asp | Ser | *U* |
| | Ileu | Thr | Asp | Ser | *C* |
| | Ileu | Thr | Lys | Arg | *A* |
| | Met | Thr | Lys | Arg | *G* |
| *G* | Val | Ala | Aspn | Gly | *U* |
| | Val | Ala | Aspn | Gly | *C* |
| | Val | Ala | Glu | Gly | *A* |
| | Val | Ala | Glu | Gly | *G* |

Figure 4.1: Francis Crick's genetic code table. Source: http://www.biologydiscussion.com/genetics/genetic-code/the-genetic-code-genetics/67689

tional. Only a few attempts have been made in biosemiotics to define the genetic code not purely relational (as a correspondence between two organic strings). These include Anton Makroš (Markoš 2002) and Alexei Sharov (Sharov 2010, 2016) who regard meaning in the genetic code as a function in the organism or behavioural activity of a creature (and not as a linear string of amino acids). Another functional approach towards the meaning was presented in works by Sungchul Ji (Ji 1985, 1999). In these attempts, meanings of the genetic code were defined not as strings of amino acids, but as functions.

Even though the elementary notion of a code and its relational understanding is shared by both semiotics and biology, function as one of the basic explanatory ideas for the most primitive molecular process in biology, or even more courageous notions like meaning or interpretations, led to misunderstandings between biologists and semioticians. Biologists and semioticians never arrived at a consensus in this regard, recalling, for instance, the famous discussion between Umberto Eco and Giorgio Prodi or the dispute between Anton Markoš and Marcello Barbieri. The discussion between Eco and Prodi culminated in 1986 on the occasion of the NATO Advanced Research Workshop on The Semiotics of Cellular Communication in the Immune System held at Ciocco, Lucca, Italy in 1986. Papers by Eco, Prodi and other semioticians and immunologists are available in the proceedings from this conference (Sercarz et al. 2013). Eco presented a rather sceptical point of view in regards with the use of term "code", yet his scepticism was not strictly categorical, since he finished his paper claiming that

> My conclusion is that your [that one of immunologists'] C space seems simpler than ours. Nevertheless it is still a space [...] Which means that in the depth of biological process lie the elementary mechanism from which semiosis springs. (Sercarz et al. 2013).

And the very last sentence: "Let me stop. I feel afraid," indicates that Eco's scepticism is related more to our profound ignorance of the very functioning of cellular processes, which troubles him, and not to an a-priori refusal of semiosis at the elementary level of cell processes. Eco's point of view on this regard is also summarized in (Eco 1984) and (Eco 1990). The reciprocal critiques between Markoš and Barbieri are illustrated in personal letters, partially published in (Markoš 2010), yet various remarks are present also in other papers, e.g. (Markoš and Faltýnek 2010; Barbieri 2011, 2012).

The arbitrariness of the genetic code is often forgotten in critiques of linguistic analogy, which is an inaccuracy that facilitates disclaiming semiotic explanations of life processes. In this manner, Eco talked about "blind material wisdom" where "elements in play are coupled together because of a stereochemical complementarity, for the same reasons (so to speak) for which a given key fits a given keyhole" (Eco 1984, p. 183). Eco's argument is a bit misleading since he speaks both about transcription and translation in the protein synthesis, attributing stereochemical complementarity to the whole process, while stereochemical complementarity plays a role only in transcription; in the translation process, on the contrary, no chemical affinity between the two elements (nucleic bases and amino acids) is present. On the other hand, Eco is probably right in his critique that when sustaining that the molecular processes at the level of protein synthesis are based on stimuli, a quite automatic "blind" execution of an instructional code which leaves no space for interpretation. We can classify this as a deontic kind of "s-code" (Eco 1984). In fact, even if we attribute arbitrariness to the genetic code, it does not yet attribute the semiotic nature of molecular processes, see also (Lacková, Matlach, and Faltýnek 2017). Arbitrariness by itself is by no means ensuring semiosis and can work also for machines, computers, etc., for example, any kind of code, "s-codes" included. In fact, a computer analogy (Monod 1972; Davidson 2010; Neuman 2008) is widely used in genetic and molecular biology.

Therefore, from a semiotic point of view, the definition of the genetic code as a relation between nucleic bases and amino acids is not satisfactory. The functional approach to the genetic code in some way contradicts the understanding of amino acids as "meanings" of genetic code. Since amino acids in the form of a string are not the final product of protein synthesis and do not represent functional units, they cannot be considered to be the meaning of the genetic code. Amino acids as such have no direct function in a cell. They only provide a framework of the final protein, which acts as functional unit, and the shape of a protein determines whether the protein can interact with other molecules and in what way. The functional definition of meaning in the genetic code has been discussed in the field of biosemiotics (Markoš 2002; Sharov 2010, 2016; Ji 1985, 1999), but its further implications have not been yet considered. In fact, if the content of the genetic code (meaning) is considered as a function of a protein (or functions inside organisms or in the biosphere in general) and not the amino acids string, then it is crucial to reconsider also the notion of expression. I propose to consider the amino acid chain as the expression of the genetic code and to exclude the level of nucleobases, since nucleobases are only secondarily related to the protein function. The order of amino acids code for the protein function and the order of nucleobases code for every single amino acid. Hence, the "nucleobases code" or DNA seems not to code for the protein function, in other words, it codes for the function only secondarily, only indirectly.

However, can an amino acid string be considered as an expression level of a code? As was said above, arbitrariness is one of main characteristics of a code (a semiotic system). Arbitrariness is also the reason why the genetic code was called "a code". Despite the fact that arbitrariness doesn't suffice as a definition for a code (as previously mentioned), arbitrariness is still a necessary condition for it.

The arbitrariness is defined as the unmotivated relation between the signifier and the signified:

> The bond between the signifier and the signified is arbitrary. Since I mean by sign the whole that results from the associating of the signifier with the signified, I can simply say: the linguistic sign is arbitrary. (de Saussure 2011, p. 67)

Does the amino acid chain satisfy the condition of arbitrariness? The particularity of relation between an amino acid string and the protein's shape is that it is direct: amino acids themselves interconnect and fold to create the protein's shape. It may seem that the relation between amino acid strings and the function of a protein is direct. Therefore, there would be no space for arbitrariness. This might also be the reason why—among semiotic approaches towards the genetic code—so little attention has been devoted to the amino acid chain. In fact, there is no doubt that since proteins are made of amino acids, there is a physical connection between those two entities. Consequently, is seems that the relation is purely physical or stereochemical. Yet, it does not mean that the relation between the amino acids and the function of the protein is not arbitrary.

I will try to demonstrate that, notwithstanding the physical connection between amino acids and protein shape, there is in fact an arbitrary relation between these two entities. To start, it is important to remember that a string of amino acids is not the only exclusive factor producing the final shape of a protein. Spatial conformation in a given context will finally define its function and for this spatial conformation to arise the protein-folding conditions are necessary to be added to the information encoded in the DNA strings. In addition, a change in the spatial conformation of a protein can change the very function of the protein, see (Kusebauch et al. 2014), for example. Proteins that are chemically identical, made of the same amino acids string, can take different shapes under the influence of other interacting molecules, thus they can have very different functions. This is the case of many protein function changes that are fundamentally rooted in protein-folding, e.g. Creutzfeld-Jakob disease (Sikorska et al. 2012).

There is not an exclusive relation between amino acids string and protein function, thus the relation between amino acids and protein function can be called arbitrary in the same way as F. de Saussure attributes arbitrariness to even onomatopoeic words. Onomatopoeias represent a special kind of words often considered as motivated by external reality (sounds in nature), hence non-arbitrary language signs, or motivated language signs. Yet F. de Saussure, in his *Course in General Linguistics* (de Saussure 2011, p. 69), notes that onomatopoeic words are as arbitrary as any other word, because even if there is a relation to external reality, the final sound is always the result of phonetic evolution and dependent on phonetic laws which themselves are arbitrary. The point is that even though there is a connection between physical reality (sounds in nature) and onomatopoeic words, physical reality motivates the linguistic sign, there is always room for arbitrariness and this is explained by the internal arbitrariness of the phonetic laws. The relation between phonetic units themselves is arbitrary: no deterministic explanation exists for the number and particular character of phonemes within a language: the Italian phonologic system has seven vowels, the Czech system has ten vowels, and the French phonological system has seventeen vowels. Still, every single one of above mentioned languages does its job to express the same meanings. And, all of the above enumerated languages have onomatopoeic words that differ even among themselves. Onomatopoeic words are motivated by an external reality; yet they are arbitrary, simply because the phonemes they are composed of are arbitrary in themselves.

This leads to the observation that the relation between signifier and signified is not the only aspect of arbitrariness. In *Course*, it is said that "arbitrary and differential are two correlative qualities," (de Saussure 2011, p. 118), which implies that the meaning is carried by the phonic differences and not by the sound itself.

> The important thing in the word is not the sound alone but the phonic differences that make it possible to distinguish this word from all others, for differences carry signification. This may seem surprising, but how indeed could the reverse be possible? Since one vocal image is no better suited than the next for what it is commissioned to express, it is evident, even a priori, that a segment of language can never in the final analysis be based on anything except its noncoincidence with the rest. Arbitrary and differential are two correlative qualities. (de Saussure 2011, p. 118)

The quote seems key to understanding the term "arbitrary" in continuity with Saussure's term "value". In this context, that is to say, in a broader context that involves the whole of Saussurian theory, it becomes clear that we must not limit the arbitrariness of the sign to the vertical relationship between the signifier and the signified, but it must be widened to the horizontal axis (differences of the signifiers between them). Hence, "differential" and "arbitrary" act as correlative qualities. A similar observation was already made by A. Martinet (Martinet 1957).

Figure 4.2: Arbitrariness on the horizontal axis: linguistic value (de Saussure 2011, p. 115).

The concept of arbitrariness is closely related to the concept of convention. If the relation between the two sets (phonologic/amino acids and semantic/protein function) is arbitrary, the only possibility to conserve this relation is convention. Take, for instance, the phonologic conventional relation between specific sounds (phonemes) and the letters of the alphabet which represent the specific sounds. It is given by conventional rules that in Czech, the sound /z/ is represented by the letter “z”. Sometimes convention might also have exceptions, for example, in the cases of assimilation, a process by which sounds become similar or identical to nearby phonemes. The similarity might be created by the placement of articulation or voiceness. In Czech, in the word “filosofie”, the “s” might be (and commonly is) pronounced as “filozofie”, in which case the letter “s” represents the sound /z/. This is an exception to conventional rule, but still the exception itself is another occurrence of convention, since the exception is conventionalised. In this case, the convention is ensured by orthography (which is an institutionalized case of conventionality). In the case of the spoken language (because there is an orthophony as another institution of language conventionalization) the conventionality is maintained by repetition of the sign usage and by mutual understanding of the communicants). In the case of biological codes, similar as the case of natural language, convention plays an important role and is enabled by evolution.

I have argued for the amino acid chain to play the role of the expression level (protein function being the content). Then, what is the role of nucleic acid strings in the proposed semiotic model? A particular string of DNA is a substantial variant of the sign (amino acid string). I already mentioned that it has the characteristics of an “s-code”, that is, a variant of a system of signs (without any connection to meaning) analogous to writing and spoken text. Written text is a variant of spoken text, and so it is an “s-code”: the written text is a transcription of a spoken text, but it is the spoken text which is connected to meaning. The written variant is only a change of substance. Also, in the genetic code, the change of an amino acid chain in the DNA is only a change in substance.

As it was mentioned above, from a semiotic viewpoint the notion of a code in molecular biology is not as evident as it seems to be. The difference lies between the semiotic and biological approaches towards the notion of code. Moreover, the distinction between them is not obvious; even semioticians who want to find a linguistic analogy in biology made the mistake of not differentiating between them. They simply took the already established definition of the genetic code from biology and analysed it from a semiotic viewpoint. Take an example from Deleuze, chapter *Geology of Morals* of the Thousand Plateus (Deleuze 1988). Deleuze talks about expressions and contents in the genetic code, but he simply takes an already existing definition of the code from biology. Consequently, he talks about the nuclebases as expression and amino acids as content.

> The essential thing is the linearity of the nucleic sequence. The real distinction between content and expression, therefore, is not simply formal. It is strictly speaking real, and passes into the molecular, without regard to order of magnitude. It is between two classes of molecules, nucleic acids of expression and proteins of content, nucleic elements or nucleotides and protein elements or amino acids. (Deleuze and Guattari 1987, p. 59)

Deleuze is fascinated by the biologic difference between “expression” and “content”, which he says is not only formal, but is real. Yet, he does not consider the very semiotic nature of the relation between nucleic acids and amino acids; he does not see that it is not a code in the strict sense, but a mere “s-code”, a substantial variant. Deleuze is right in that the distinction is not formal, because in a matter of a fact, the form does not change, as it does in cases of exchanging one “s-code” for another “s-code”. For instance, if one transcribes a spoken text into a written text, then translates a written text into Morse code, then translates the Morse code into sign language. All those are “s-codes” related to the same content level. The distinction is purely substantial, the form remains the same. Thus, the distinction is obviously not formal. However, the distinction between expression and content should perhaps be formal, expression and content being of two different orders: one linear and the other non-linear.

Deleuze’s analysis is too closely attached to classical definitions of organic processes which are found in biology textbooks. He is accenting the linearity of the nucleobases string as leading to the final dimensional protein, implying the direction “from sequence to structure”, which is taken from biology.

Deleuze's work is interesting in that he attributes relativity to content and expression, he suggests that expression is always to be understood in connection with a given meaning and vice versa, and content is relative in that it may become an expression for other content. This presupposes that there is more than one kind of expression and content level within organic strata:

> Proteins of content have two forms, one of which (the infolded fibre) plays the role of functional expression in relation to the other. The same goes for the nucleic acids of expression: double articulations cause certain formal and substantial elements to play the role of content in relation to others; not only does the half of the chain that is reproduced become a content, but the reconstituted chain itself becomes a content in relation to the "messenger". (Deleuze and Guattari 1987, p. 59)

Not even Deleuze—and this is quite paradoxical since his essay comprises part of the famous theory of the rhizome (a semiotic space which is pluridirectional)—escaped from the imposed unidirectional order "from sequence to structure". Even if expression and content are relative, they are still slaves of the imposed order. All the more, the double articulation is a slave of the imposed order:

> The cellular chemistry presiding over the constitution of proteins also operates by double articulation. This double articulation is internal to the molecular; it is the articulation between small and large molecules, a segmentarity by successive modifications and polymerization. First, the elements taken from the medium are combined through a series of transformations…All this activity involves hundreds of chemical reactions. But ultimately, it produces a limited number of small compounds, a few dozen at most. In the second stage of cellular chemistry, the small molecules are assembled to produce larger ones. It is the polymerization of units linked end-to-end that forms the characteristic chains of macromolecules. (Deleuze and Guattari 1987, p. 42)

Here Deleuze not only followed the imposed order of Modern Synthesis (the notion of Modern Synthesis will be explained in the next section) in biology, but he also fell into the linguistic trap of treating language phenomena as a generative mechanism which combines expression units into content units. The order imposed by linguistics goes as follows: phonemes (expression) bind sequentially and form words (content), words bind sequentially and "fold" by flexion[10] (expression) and form sentences (content). Analogically, according to Deleuze, DNA molecules bind sequentially and by combinatory rules form RNA molecules, RNA molecules bind sequentially and form protein fibre, protein fibre folds and forms proteins. He describes double articulation in this manner. Unfortunately, Deleuze is wrong in that double articulation should not been understood in this direction, at least not as A. Martinet (Martinet 1967) designed it to be: the first articulation divides a sound chain into units of meaning (words or morphemes), the second articulation concerns further subdivision of single meaning units into meaningless elements (phonemes). The result of the first articulation is an open set of elements; the result of the second articulation is, on the other hand, a closed set of units.

Maybe it is difficult, and in some way unintuitive, to consider organic strata without imposed order, which would mean to liberate our understanding of organic strata from directionality sequence-structure. As I suggested above, there are two principal origins of this unidirectionality.

(I) The first one is obvious, and comes from basic biology as it is taught even at the high school level: protein synthesis goes from nucleic acids to proteins and is unidirectional and irreversible. This is the so-called dogma of Modern Synthesis. It has to be remarked here that this dogma is being reconsider thanks to recent discoveries in epigenetics, unidirectionality is not so strict any longer and is the question of current discussion in epigenetics and evo-devo theories (I will comment on this factors below).

---

[10] In Czech, the term "folding" is used both for syntax (větná *skladba*) and protein folding (*skládání* proteínů).

(II) The second origin of unidirectional thinking comes from linguistics. Since the discovery of the genetic code, biology has been connected to linguistic terminology and linguistic analogies. It was Roman Jakobson who proposed the most commonly used DNA-language analogy, which analogises of DNA bases with letters, then triplets of bases with words, and finally genes with phrases (Jakobson 1971a, pp. 655–696). Jakobson suggested that speaking directly about the language-like structure of the genetic code, which, according to him, consists of letters (DNA bases), words (codons) and sentences (genes). He consequently pointed out that features, such as synonymy, suprasegmental or syntactic delimitation, the system of distinctive features and static plasticity / flexible stability[11], are found in the genetic code. Jakobson's linguistic interpretation of genetic code was later further elaborated, e.g. (G. Katz 2008) and the Jakobsonian linguistic metaphor is basically the most common understanding of the genetic code as such. The invalidity of Jakobson's analogy was proven by quantitative methods (Faltýnek and Matlach 2016). Notwithstanding the utility of linguistic metaphors for pedagogical purposes, it has some significant imperfections. Linguistics, or at least linguistics as presented in Jakobson's metaphor, are strictly unidirectional and only study expressions. Or, better, study the way how expressions generate contents (from phonemes/letters to words and sentences). Very few linguistic approaches consider the opposite situation, that is, to study how contents are imprisoned in the expressions. The reason why content-based linguistic approaches are rather uncommon and unappreciated will be discussed in detail in the last chapter of this thesis. For the moment, the only observation I want make here is that the classical linguistic expression-based approach which Jakobson applied to biology led to a somehow limited understanding of biological phenomena in semiotics (from Eco do Deleuze).

---

[11] A term proposed by Czech linguist Vilém Mathesius (Mathesius 1983).

## 4.2 Extended Synthesis

As shown previously, semioticians often call into question the semiotic nature of the biological process primarily due to the blind automatized nature of the genetic code, that is, the code between triplets of nucleic acids and the amino acid chain. They are in the right. At the level of nucleic acids, the genetic process is quite machine-like and the metaphor of a computer seems more appropriate than the metaphor of a conscious reader or interpreter. The process of transcription is actually based on a chemical affinity and there is no space for semiosis or interpretation; it is a pure chemical phenomena. At the level of translation, arbitrariness already comes into play. There is no chemical affinity between nucleic acids and amino acids, yet this process is still machine-like and could be called "deontic institutional s-code" because it has mostly "if x then a y" nature. The critiques by U. Eco were legitimate. The genetic code is a code, but it is not a matter of semiosis. This is the picture the classical Modern Synthesis created and that this is the picture limiting semiotic approaches towards the molecular processes because, by its nature, the Modern Synthesis dogma of molecular biology excludes any interpretation space and potentiality: the direction of protein synthesis goes from genes to proteins and is an automatic and unidirectional process.

In recent years, however, the very notion of gene is being discussed in the field of genetics in terms of base of function (Scherrer and Jost 2007) points out that the notion of gene is not particularly clear in that the gene as an expression of a certain function is in many cases not fully present in the level of DNA. Instead, regulatory aspects are as important as the information encoded in the DNA. With this point of view, the definition of gene as encoded in the nucleic bases strings is unsatisfactory.

Recent findings in life sciences have led the scientific community to discussions and revaluations, to the point that even some of the basic genetic concepts are losing their solid foundations. The very notion of gene has been questioned, with the discussion based on speculation as to whether we understand gene as having a heritable function or rather in terms of its localisation in DNA molecules. It is known from recent discoveries that these two understandings do not always coincide, for more information on this topic see (Snyder 2003; Scherrer and Jost 2007). Finding a solution among these disagreements is a concern not only to biologists, but also to philosophers and semioticians, when one considers that the major part of the problem is conceptual rather than empirical:

> Specifically, the question „how much of the (human) genome has an identifiable function" is discussed controversially. Estimates range roughly from 5 to 90 % [...] Such divergence cannot be reconciled by more accurate data. Rather it reflects dramatic disagreements about the proper definition of function. (Laubichler et al. 2015, p. 144)

As the definition of gene is a current matter of discussion in the philosophy of biology, the definition of the genetic code may also be reconsidered. Epigenetic modifications, epigenetic inheritance, gene expression regulation and many other recently discovered molecular processes may reopen the discussion about the accuracy of the semiotic approach in biology. Extended Synthesis in some ways answers questions that are quite intuitive, but which the DNA-centralized genetic theory of Modern Synthesis cannot withstand because of negating potentiality and non-unequivocal code usage. If every cell in our body has an identical DNA script, what makes cells of special body parts so different (for example, liver cells and skin cells)? How can species diversity be explained, if we consider, for example, that humans share 99% of their DNA with chimpanzees and 92% of their DNA with mice? The answer is that the discovery of the regulatory function of the non-coding sequences of DNA ("junk" DNA) offers is simply that even if the cells have an identical DNA, not all the information stored in DNA strings is expressed. This means not all genes are transformed into proteins and it is the non-coding DNA that provides instructions to create regulatory units that decide which genes are going to be expressed and which ones are going to be silenced. These regulatory activities may be influenced by the external environment: if the environmental conditions are not appropriate to a particular trait, the genes that code this trait will be silenced by regulatory units and vice versa.

It should be noticed here that the very discovery of regulatory DNA was preceded by an application of linguistic methods to DNA strings by Mantegna et al. (Mantegna et al. 1995), see also (Havlin and al 2003), who analysed manifestations of Zipf's law in coding and non-coding DNA. Zipf's law is one of the most importat laws of mathematic linguistis (Zipf 1949, pp. 22–25). Coding DNA, according to Mantegna et al., exhibits Zipf's law. Noncoding DNA manifests Zipf's law as well, but only to certain extent. Mantegna's analysis was motivated by the finding that only a small portion of genome codes (considered 5.33% in Homo sapiens) carries information for the construction of proteins. The remaining part of the genome does not have such a clear function and in the 1960s, the term "junk DNA" was used. It was believed that non-coding DNA in the genome is a historic relic without any use to the organism, consequently it was called junk DNA or "silent DNA". Mantegna et al. (Mantegna et al. 1995, p. 2949) further argued that non-coding DNA resembles some natural language properties. Mantegna discussed the fact that non-coding DNA, in terms of redundancy, exhibits even more natural languages properties, in comparison with coding DNA. Mantegna thus extended the analogy of DNA and natural language, and again applied the linguistic metaphor to DNA. These findings led to the hypothesis that non-coding DNA also has a function which is currently unknown to us. Later development in molecular biology confirmed Mantegna's hypothesis and found that this function is to regulate gene expression, for example (Consortium 2012, p. 57).

The new discoveries argue that not only the virtual script, but also its context-dependent reading, determines the final result (cell, protein or animal) and that these expressions of the script can change due to environmental conditions and the changes can be reversible, but in some cases can also be heritable. There is already considerable empirical evidence of epigenetic findings. The research is based on demonstrating several epigenetic processes, e.g. how non-genetic inheritance is possible, how the gene-regulation functions, how the external factors (environment) influence the gene expression and what is the exact role of non-coding sequences in DNA macromolecules. Among these experiments, research in lateral (horizontal) gene transfer also appears to have an important influence on discrediting “the dogma” of the Modern Synthesis. Lateral gene transfer, present primarily in bacteria, attacks the core idea of Darwin’s tree of life and linear species evolution. Considering lateral gene transfer, not a tree-like but rather a web-like theory of evolution was proposed by John Dupré.

> The Tree of Life is the standard neo-Darwinian representation of the relatedness of organisms. As a tree, crucially, it constantly branches, and branches always diverge, never merge. Species are represented as small twigs, larger branches represent larger groups of organisms. But this image of the Tree of Life has been rendered at least partially obsolete by recent developments, especially in microbiology, where so-called lateral gene transfer, the passage of genetic material not from ancestors, but from sometimes distantly related organisms on widely separated branches of the Tree of Life, is common. When we note that if lateral gene transfer is common, the overall structure of relations between organisms will take the form not of a tree, but of a web, or net. And in a web, unlike a tree, there are many paths from one point to another. (**dupré2012**).

This evolutionary model extends the one-dimensional species evolution to at least a two-dimensional evolution (classical genetic downward inheritance and lateral gene transfer). Dupré focused on lateral gene transfer in bacteria, but it is possible to enlarge Dupré´s web model with additional molecular processes, not necessarily genetic. If one also considers reversible epigenetic processes, one obtains not only two dimensions of evolution (vertical and horizontal), but also two directions (forward and backward) and the model becomes a genuine web within which passages between nodes in every direction are possible.

Extended Synthesis replaced the tree-like model of evolution with a web-like model of evolution, allowing in a similar manner, reconsidering of the notion of code. Considering multidimensional gene transferring, the actual definition of genetic code as one-dimensional and one- directional (from DNA to proteins) has to be reevaluated as well. If the role of the genetic code is supposed to be the encoding of heritable traits, the genetic code as presented by Modern Synthesis might not be up to this work in the context of recent findings (e.g. epigenetic modifications or lateral gene transfer). Additionally, since the notion of code by its nature is too narrow in light of Extended Synthesis, relating only two distinct worlds, perhaps another notion may embrace the entire scale of phenomena better. Umberto Eco in his later works replaced the term code with the term "encyclopedia". I propose adopting an analogical shift in biosemiotical thinking and reconsidering the notion of code in terms of an encyclopedia-like model.

Umberto Eco elaborated his theory of encyclopedia, in opposition to dictionary-like representations of semiotic content (semantics), as a critique of the Porphyrian tree (Eco 1984, 2007). A dictionary-like representation of content is based on hierarchy of semantic units, graphically illustrated in a form of a branched tree, with a finite set of universals- basic units in positions of the last branches of the tree. These basic units are also called semantic primitives. The problem of dictionary-like representation is that it presupposes semantic content being definable by a finite inventory of basic units, yet it does not define basic units themselves and does not explain how to arrive at the desired inventory of ultimate semantic primitives. An encyclopedia-like representation, in contrast, assumes that the representation of the content takes place only by means of interpretants, in a process of unlimited semiosis. These interpretants being in their own interpretable. There is no bidimensional tree able to represent the global semantic competence of a given culture (Eco 1984, p. 68).

A dictionary-like model of evolution (the Darwinian tree of life), similarly to the semiotic content theory, presupposes a finite set of basic elements, which, by combinatory operations, create the global space of biodiversity that is in its turn exhaustively definable only by the mere catalogue of the basic units. The metaphor of dictionary in the field of genetics was already proposed by F. Crick (Crick 1982, p. 171).

Crick described the genetic code as a “small dictionary” of correlations between the language of 4 letters of genetic material and the language of 20 letters of proteins. Accurately, for the translation process, a dictionary is needed. The dictionary metaphor is a very elegant solution, explaining the entire phenomenon of biodiversity by combinatorics of merely four basic elements. It has turned out, however, with the most recent discoveries in life sciences, that things are not that simple. According to Extended Synthesis, there is, apart from the genetic code, a vast range of phenomena that are needed to explain processes in living creatures and that may help reach a deeper understanding of the evolution of species. Epigenetics, for instance, accounts for features that, if considering only and exclusively four letters of the genetic dictionary, would remain unexplained: epigenetic modifications can be chemically expressed by the addition of a special mark, e.g. adding a methyl group to the DNA molecule so that one of the four letters of the genetic dictionary becomes a marked letter, cytosine becomes methylated cytosine, this modification serves many cell processes such as aging or inactivation of the X chromosome. A number of studies have been conducted concerning maternal care in rats (Weaver et al. 2004; Cameron et al. 2008). In these studies, maternal licking and grooming was identified as an important factor in the later adult sexual behaviour of the offspring. Two kinds of rat mothers were studied, high-LG (licking and grooming) mothers and low-LG mothers. The results suggest that female offspring of low-LG mothers show an increase in sexual receptivity (e.g. vaginal opening appears significantly earlier in life than in the case of the offspring of high-LG mothers). These variations in rat behaviour are not random and are observable at the chemical level by the DNA methylation, which functions as “diacritical” changes at the DNA script.

Another example of epigenetic modifications are the particular butterfly wing colour-patterns produced physiologically in response to environmental stress, e.g. temperature conditions (Hiyama, Taira, and Otaki 2012). As with many other epimutations, the particular wing colours and patterns can become heritable and transmitted to the next generation.

DNA methylation is only one example of the entire scale of adding letters to the four-letter genetic dictionary. The finite number of "genetic universals" consequently comes to question. What can be accomplished at this moment, having an inventory of epigenetic "diacritic" marks, is simply to enlarge the dictionary by extending it with new marked letters. This proposal would though not resolve the problem, since Extended Synthesis accounts for the influence of environmental stress and interaction with other organisms as actively participating in creating new marks. Since changes in the environment are unpredictable, extending of the dictionary model seems inappropriate: it will never encompass exhaustively all possible new formed epimutations. Instead, another model appears plausible: a model that allows the internal dynamics of the system with the impossibility of a finite set of universals, but with the possibility to interact between elements. Such a model was proposed by Umberto Eco and was called encyclopedia. Already anticipated by Laubichler, the web-like model of evolution corresponds to the basic characteristics of a semiotic encyclopedia, also illustrated as a net.

> The main feature of a net is that every point can be connected with every other point, and, where the connections are not yet designed, they are, however, conceivable and designable. A net is an unlimited territory. A net is not a tree. The territory of the United States does not oblige anybody to reach Dallas from New York by passing through St. Louis, Missouri: one can also pass through New Orleans. (Eco 1984, p. 81)

The comparison of the passage from Dallas to New York can be analogized in biological terms as follows. There is no unique way to reach a certain state of an organism. Sharing and transmitting of genes is possible in contact and also in a distant way in both horizontal and vertical dimensions. By contact transmission, I mean direct, for instance, parental gene transmission from parent to offspring. By distant transmission, I mean the processes which do not require physical contact between the cell donating the DNA and the cell receiving the DNA (e.g. bacterial transduction).

| | Contact | Distant |
|---|---|---|
| Horizontal | Parents-offspring | Horizontal gene transfer |
| Vertical | Cross-over,<br>Endosymbiosis,<br>Bacterial conjugation | Vertical gene transfer<br>(Transduction,<br>Transformation) |

Table 4.2: Various ways of gene transmissions. Breaking down the one-dimensional model of evolution.

In this context, it is worthwhile to mention a work on nucleotide codes by Edward Trifonov. Trifonov elaborated the theory of the so-called "Gnomic" (Trifonov 1988), a language of nucleotide sequences that encompass several codes within the passage from DNA to proteins, the "genetic code" (triplet code of amino acids and nucleotides) being only one part of "Gnomic language". According to Trifonov, "Gnomic" is a multicode language that even enables an overlapping between the codes. The codes Trifonov speaks of are the DNA replication code, the DNA-to-RNA transcription code, the chromatin code, the RNA-to-protein translation code and the RNA-to-DNA reverse transcription code. Trifonov's theory represents an innovative understanding of genetic code and its functioning, detracting attention from DNA triplets as unique elements having the power to determine all genetic information. Further elaborated in the 1980s, the theory of "Gnomic" language introduced new approaches in biosemiotics.

Trifonov's theory may appear similar, at first sight, to the encyclopedic model, considering the multiplicity and overlapping of codes. There is one important difference, however, in the two models. Trifonov strictly separates codes from what they refer to. To put it more clearly, he only describes "s-codes", since he speaks of "codes" without considering them as bridging between two independent worlds, but considering them as a set of internal syntactic rules. This observation becomes more evident with the following quotation:

> "Both Gnomic and human languages can be expressed in the form of one-dimensional arrays (texts) of symbols, elements of limited alphabets. Both languages reflect,

> however, the four-dimensional (space and time) reality, though of different kinds." (Trifonov 1988, p. 510)

The conjunction "however" implies that there is a contradiction between the two propositions. Trifonov in this way separates the written (one-dimensional) form of an "s-code" from what it reflects (four dimensions), putting them in contradictory relation. He excludes one from the other so that in his understanding, the notion "code" is only related to the (one-dimensional) expression, not to the content (four-dimensional).

The encyclopedic model, in contrast, encompasses both expression and content, since a given content may become an expression in another sign relation and the process of unlimited semiosis goes ad infinitum. In a similar manner, an organism (being the content of a given genetic script) becomes in its turn an expression in influencing the genetic script by epigenetic modifications for a new content (its offspring). The two worlds related by a code make both part of the encyclopedia.

In summary, Extended Synthesis disclaims the traditional understanding of genetic code as a univocal and unambiguous relation between two linear sequences: nucleobases sequence and amino acids sequence. The notion of code, in the light of Extended Synthesis, loses its basic characteristic of unidirectionality, with the pluridirectional notion of enyclopedia being better suited. The model of encyclopedia was elaborated as a critique of the tree-like model, yet this does not imply that the tree (dictionary) should be excluded from the encyclopedic model. On the contrary, it makes up part of it. Eco even suggested using a dictionary as a suitable tool for a certain type of analysis (Eco 1984, pp. 84–86). In other words, the exclusion of the dictionary-like model is not tenable for encyclopedia theory, nor it is tenable for Extended Synthesis in biology, which may perfectly match with particular pieces of encyclopedia. In the next few paragraphs, I will illustrate the ways in which a dictionary makes up part of an encyclopedia in a biological context.

## 4.3 Habit as Final Interpretant

In the semiotic theory of C. S. Peirce which U. Eco was inspired by, the process of semiosis is potentially unlimited. This understanding of semiosis as an unlimited interpretation, with the Interpretants themselves being interpreted, represents a driving mechanism of Eco's semiotic encyclopedia. Peirce also introduced the concept of Final Interpretant, allowing for a kind of ultimate interpretation of a given sign. The Final Interpretant is related to the notion of habit, with habit or convention playing a crucial role in Peirce's pragmatism. It allows for quickly reaching a consensus on reality in a given context. Force of habit temporarily freezes the infinite interpretation process, the unlimited semiosis.

> The Final Interpretant does not consist in the way in which any mind does act but in the way in which every mind would act. That is, it consists in a truth which might be expressed in a conditional proposition of this type: If so and so were to happen to any mind this sign would determine that mind to such and such conduct. (Peirce CP, § 8.315)

The Final Interpretant is a stabilized interpretation of a given sign that, by force of habit, renders interpretation a kind of automatic process of meaning attributing. This is, in the pragmatic point of view, a very useful device that facilitates actions in the world, inasmuch as an individual or a culture (or species) occurs in exactly the same or analogical situations in everyday life. The automatic process of living cells may also be seen hereby as a result of habit-making that was not always an automatic process. A model of evolution based on habit-making was already proposed by K. Kull.

> We can formalise behavioural habits—which always are acquired rules—as functional links, as operations, or algorithms. Interpretation—as much as it leaves traces—designs links, which turn, stepwise, into habits, and algorithms; these links may become parts of an organism's structure. (Kull 2015, p. 227)

One may understand automatic cell processes, stimuli in animal behaviour and also the genetic code as a result of a habit, which was preceded by decision making and interpretation. Moreover, one should bear in mind that the Final Interpretant is finally only temporary and locally and it can be, in its turn, interpreted in a new way by the influence of a different context. In this way the unlimited semiosis recommences and leaves open space for interpretations. The Final Interpretant freezes the process of semiosis only temporarily, and at the point when the Final Interpretant gives birth to a new interpretation, the semiosis begins once again. Modern Synthesis introduced the notion of the genetic code as a blind correspondence between elements. It was not supposed to become the object of an interpretation, the semiosis was not considered to exist. With the arrival of Extended Synthesis, the genetic code became only one part of a complex web of processes responsible for the gene expression and for heritability of traits, with automatic matching of DNA triplets to amino acids playing a crucial, yet indefinite and unique role, always leaving the possibility of "waking up" the frozen Final Interpretant and starting the semiosis.

Extended Synthesis embodies a large scale of various phenomena that combine amongst themselves to make part of an encyclopedia. There is no room in this thesis to comment on every one of them and it would be impossible to describe exhaustively in which manner every piece of encyclopedia is connected to other pieces, if we consider that the possibilities of interacting and connecting distant entities are potentially infinite. As U. Eco pointed out, encyclopedia as a whole is inexpressible and the encyclopedic representation is never global, yet is always local (Eco 2007, p. 57). Hence, we can make "a local cut" within the web of the biological encyclopedia to see one particular level, to illustrate how a specific encyclopaedic level may constitute a local code, that is, in what manner two independent worlds can correspond with one another. An example of such two distinct worlds will be taken from genotype-phenotype correspondence. Modern Synthesis is a concept that lacks further clarifications on how phenotypical variations are possible without change in the DNA script, or vice versa, how it is possible that a given phenotype may be obtained by different DNA scripts. Modern Synthesis' answer to the question as to what is the relation between genes and their actual expression is that phenotypes are simply blueprints of genes, this relation being understood as an unequivocal correspondence. Recent studies in evolutionary biology, genomics and proteomics and other life sciences have yielded, however, the untenability of this simplistic, yet elegant model. Adopting a semiotic point of view, having a gene and its expression as elements taken from two independent worlds, the former being a sign and the latter its object, an interpretant that guarantees the passage from one to another is missing. Studies on evolvability and genotype → phenotype mapping (Alberch 1991; Pigliucci 2010) are conceived as occupying the missing element, the missing transitional link between genes and their expressions.

As Massimo Pigliucci noted,

> the undeniable progress we have made in understanding G → P maps, both empirically and theoretically, is such that one should hope that evolutionary biology has reached the point of forever being past simplistic ideas like genetic programmes and blueprints, embracing instead a more nuanced understanding of the complexity and variety of life. (Pigliucci 2010, p. 564)

In another place, Pigliucci explains the importance of the concept of genotype → phenotype mapping and the need for abandoning the gene-centralised model:

> Genomics and what I refer to as "postgenomics" (proteomics, metabolomics, etc.) started out squarely within the conceptual framework of the rather gene-centric MS , with the view that once we "decode" the genome of an organism we somehow gain a universal key to understanding its biology. The reality of organismal complexity has shattered such simplistic visions [...] The complexity of the genotype → phenotype map cannot be understood only by bottom-up approaches such as those that focus on gene networks and regulatory evolution, however. Accordingly, much discussion has

> been generated around a group of tightly related concepts such as modularity (the degree of interconnectedness of different traits), evolvability (the ability of living systems to further their own evolution), and robustness (the resilience of biological systems to perturbation). (Pigliucci 2009, p. 223)

Genotype → phenotype maps represent the passage from genes to their expression in a very complex and non-unequivocal way. This concept has the potential to reinforce the semiotic explanations of life processes, replacing the genotype → phenotype unidirectional and explicit relation by a model of a pluridirectional map providing many possible transitory operations. The notion of interpretant can thus be introduced and an encyclopedic model can be applied.

# 5

# Linearity and Non-linearity

## 5.1 Folding as Syntax

In the previous section, I made use of a comparison between protein folding and the syntax of natural language. I defined both syntax and protein folding as a kind of "s-code", a coherent set of rules which, only when connected to another "s-code" (content level), leads to meanings and became a real code. As a point of fact, syntactic metaphor in biology is not my own speculation. In an effort to understand better the mechanism and rules for protein folding. Biologists resort to the linguistic analogy since the discovery of the genetic code. Mention has already been made of the notorious DNA-language analogy by Jakobson. An entire range of linguistic terminology also pervades biology, e.g. translation, transcription of DNA, reading and interpreting of the DNA script, the notion of code itself, etc. In recent studies on protein folding, the folding process is explained by analogy with syntax in natural language. The syntactic analogy (Chomskian syntax is mostly used) has appeared in numerous scientific papers on protein studies in recent years (Gimona 2006; Loose et al. 2006; Kister 2015).

It is quite easy to understand what led biologists to usage of a syntactic metaphor. Folding of proteins is a process involving connecting smaller units into larger and more complex units. In syntax, clauses or sentences (complex units) are described as complexes of words (smaller units). This is too simplistic a point of view, however, for syntax is not a mere combining of words into sentences, a lego-like building mechanism. The celebrated example by Chomsky (Chomsky 1957) demonstrates that syntactic rules allow for the creation of sentences that are meaningless: "Colorless green ideas sleep furiously". The classic Chomskean approach argues that syntax is completely independent from semantics, in other words, to study syntax, we do not have to understand the meaning of the words that constitute a sentence. The analogy of a simplistic understanding of syntax in protein folding will not help, however, because researchers are interested in functional proteins, not in any kind of possible protein. "Colorless green furiously–like" proteins are non-sense proteins that in all probability, given the chemical properties of amino acid residues, can be potentially obtained, but would be of no function (no meaning) for metabolism.

Syntax is not a mere building mechanism involving connecting smaller units into larger units. A sentence must be meaningful, within a sentence, smaller units are related together in a way to constitute a coherent and meaningful whole. The packing of the complex meaning into small units lined up together is, however, an extremely complicated process that grammarians and linguists try to understand. The great question of grammars is to resolve the relationship between sound and meaning, between linear sound and non-linear meaning. How is it possible that a linear sequence of sounds (or linear sequence of written symbols) represents a non-linear meaning?

## 5.2 Linearity

Let us first have a closer look at the very notion of linearity[12]. In the preceding sections, I used it synonymously to notion sequentiality. If, however, we study linearity as defined by Saussure (de Saussure 2011, p. 70), as one of the two main language principles (among arbitrariness), the entire thing becomes more complex. As Faltýnek (Faltýnek 2011) points out, the essence of the Saussurean term "linearity" consists of the impossibility of superposing (pronouncing) two signs at the same time. Sequentiality is, then, only an epiphenomenon of linearity. Linearity as a quality of an utterance is determined by the impossibility of superimposing two series of distinctive features, which is why signs have to be ranged one by one, and sequentiality is only an epiphenomenon, not a necessary condition (Faltýnek 2011, p. 64). Linearity is the very nature of a linguistic substance (sound or written characters) which consequently has the impossibility of superimposing more than one specific element at the same time (de Saussure 2011, pp. 70, 123). In other words, it is an example of the theorem of the excluded third. This is exactly what Peirce's theory of continuum disproved. Linearity of speech is represented by the pronunciation of one after another phonemes sequentially in time. The linearity of the written characters can be represented by left to right writing (Roman character, Russian alphabet), right to left writing (Arabic, Hebrew) or top to down writing systems (Chinese, Korean). As a matter of fact, linearity is nothing but dyadic relations and what is important to notice is that Saussure attributes it exclusively to the expressions (signifier); the term "linear nature of the signifier" (and not signified) is used. I am of the opinion that this remark is of no small consequence. Saussure, contrary to other structural theories, did not consider the content level (signifier) of an exclusively binary (dyadic) character. And this is not everything. He also argues that the idea of the non-primary-dyadic character of the signifier. He states that, before the sound chain gets in touch with meaning, it is nothing but "a continuous ribbon along

---

[12] I am not using the term "linear" in a mathematical sense, I am using it in a way as it was defined by F. de Saussure (de Saussure 2011).

which the ear perceives no self-sufficient and clear-cut division". Saussure does not use the notion of folding of continuity, yet speaks about the continuous line of signifiers that only become articulated (divided) when meanings are called into play.

> We know that the main characteristic of the sound-chain is that it is linear (see p. 70). Considered by itself, it is only a line, a continuous ribbon along which the ear perceives no self-sufficient and clear-cut division; to divide the chain, we must call in meanings. When we hear an unfamiliar language, we are at a loss to say how the succession of sounds should be analyzed, for analysis is impossible if only the phonic side of the linguistic phenomenon is considered. But when we know the meaning and function that must be attributed to each part of the chain, we see the parts detach themselves from each other and the shapeless ribbon break into segments. (de Saussure 2011, pp. 103-104)

Saussure seems to corroborate the previously presented model inspired by Peirce concerning the undecidability of the expression level unless it gets in contact with the content level: generating of dyadicity only from the triadicity. Only when we know the meaning and function, the shapeless ribbon breaks into segments. The point of De Saussure here is that the meaning determines (or generates) the form of the expression level and not vice versa. Consequently, the celebrated Saussurean claim that language is a form rather than a substance acquires an additional value, which consists in fact that the form of language does not exist pre-meaningly, it is not an entity of an individual and autonomous existence and is not in any fashion a pre-existing platonic form. To divide a line, we must call in meanings. This means that Saussure differentiates between a continuous unarticulated line on the one hand and a segmented discontinuous line on the other. An unarticulated continuous line is exactly what I had in mind in my thought experiment with protein fibre. Saussure's remark clearly demonstrates he was aware of the problem of direction of determination of the expression form. Still, this truly inspiring concept of Saussure was not further elaborated in linguistic practice. Academic grammars always treat language in the opposite direction, starting from phonology and moving towards morphology and syntax. Semantics in most cases does not make up part of the grammar (at least not in the official academic grammars).

An obvious example of this fact is the current situation in linguistics. Mainstream linguistics of today deals practically exclusively with syntax (generative syntax), which is a study of one "s-code", an "s-code" of syntactic relations which are detached from meanings: "grammar is autonomous and independent of meaning" (Chomsky 1957, p. 17). Yet, syntactic practice has demonstrated that it is impossible to deal with syntax without considering semantics. This is in all probability because, as De Saussure has argued, we must call in meanings to segment the sound chain. De Saussure formulated the need for meanings in the generating of forms, yet he did not go far enough. He did not comment on the very problem of non-linearity of meaning, but only stated that linearity is a matter of the signifier. He did not comment on the non-linearity of the signified, and did not discuss how this non-linearity is resolved in the linear sound chain. This might be one of the reasons for the shift in the linguistic paradigm from De Saussure and structuralism to Chomsky and generative syntax. Most structural linguistics[13] did not study syntax (with the exception of Tesnière, e.g.), which might be rightly considered a serious disadvantage, since the role of syntax is to resolve the mystery of passage from non-linear meaning to linear expression. When considering syntactic representations and the syntactic representation of a sentence, whether it is taken from dependency or generative grammar, it is always non-linear. This is why it helps us see the structure of the sentence, because the meaning of a sentence (and meaning in general) is not linear.

An example of syntactic representation may be taken from valency syntax. Take, for example, the verb "to give": to give something is an action that is impossible to comprehend linearly, as a sequence of A

---

[13] I am referring to European structuralism mostly. American structuralism payed attention also to syntax.

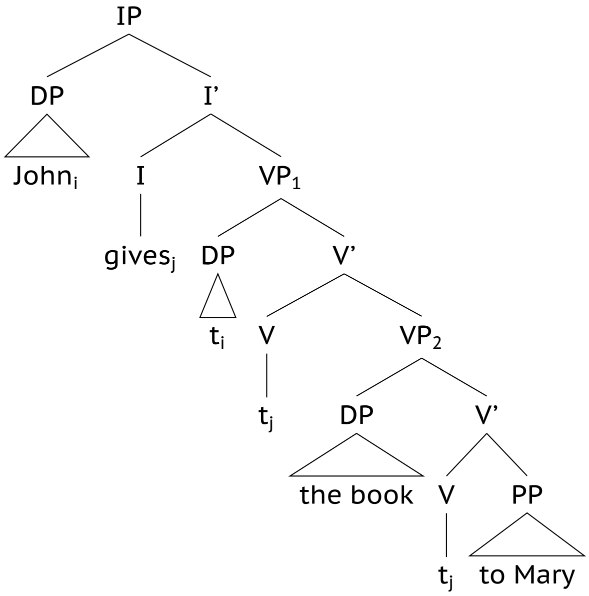

Figure 5.1: Syntactic representation of the sentence "John gives a book to Mary". Inspired by (Larson 1988).

gave B to C ⟶ C received B from A ⟶ B as given to C, the action of giving is simultaneous, for this reason the structural syntactic representation of verb valency is two dimensional and non-linear.

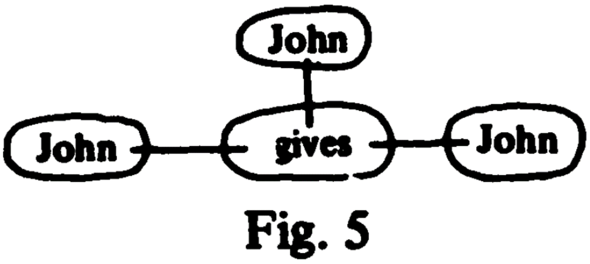

Figure 5.2: Two-dimensional representation of verb valency (Peirce CP, § 3.469)

In natural language, linear sequence does not explain the meaning of the sentence "John gives John to John"; only structural representation of a non-linear character may explain it (a two dimensional verb valence model). Linearity is an obstacle of the material world of impossibility to superimpose two or more elements simultaneously, yet is not a constitutive or explanatory principle of meaning-making in linguistics and of function-generating in proteins. The theory of verb valency is usually connected with structural linguistics and especially with Lucien Tesnière's book *Elements de Syntaxe Structurale* (Tesnière 1959). Very similar observations, if not identical ones, were already made several decades before Tesnière by Ch. S. Peirce (Paolucci 2006, 2010) and Peirce, exactly like Tesnière, used chemical valency as a model to represent verb valency.

The impossibility of the linear representation of the meaning of the sentence was even recently elaborated on a higher level, passing from a two dimensionality of syntactic trees to "a 3-D" model of folded syntactic spaces. This approach was called the topological approach.

> In this topological view, the phrase marker deforms and folds, rather than having elements move: in this way, there is no sorting mechanism required, because no reordering takes place. (Krivochen and Saddy 2016, p. 37)

Martin and Uriagereka (R. Martin and Uriagereka 2014) applied this topological "knot" model to the specific case of repetition in syntax, which can be exemplified by the sentence "Students believe that students were criticized." Martin and Uriagereka also recur to the metaphor of origami paper, already mentioned in this thesis as used by Paolucci and Deleuze.

> Imagine two sheets of origami paper, one black and the other grey, combined by stapling the bottom edge of the grey sheet to the top edge of the black sheet. We could continue this derivation by taking a new sheet of grey paper and stapling it to black sheet, so that now we have a total three sheets combined into a single object with the bottom-to-top order grey-black-grey. However, suppose that rather than introducing a new grey sheet, we instead fold the sheets from the initial step in such a way that we staple the bottom edge of the black sheet to the top edge of the grey sheet. In this case, the resulting object consists of just two sheets of paper, yet it is topologically more complex than the previous object with three sheets. Needless to say, there is something very similar about the nature of these origami operations, but at the same time we have

> created very different sorts of objects in the two cases. (R. Martin and Uriagereka 2014, p. 175)[14]

In the previous paragraphs, I tried to argue for a need for meaning-sound (or semantic-phonological) interface in order to construe a grammatical model. This need is given by the nature of language characterised by the asymmetry between the linearity of sound and the non-linearity of meaning. Krivochen also talks about an essential discontinuity that configures a tension between the sound and the meaning.

> The interaction between a procedure that increases dimensionality of syntactic structure and the need to flatten that structure dynamically yields what is known in physics as a dynamical frustration: an irreconcilable tension between opposing tendencies which gives rise to a dynamical (meta-) stability. We follow (R. Martin and Uriagereka 2014), (Krivochen and Saddy 2016) [...] in saying that there is a dynamical frustration at the core of language in the interaction between sound and meaning (see also (Tesnière 1959, p. 21)), which shapes the properties of linguistic computation and yields "mixed computation" as a natural consequence of considering language to be a dynamical, nonlinear system subject to orthogonal requirements from semantics and phonology. (Krivochen 2016, pp. 31-32)

The nature of this tension in natural language was first described explicitly by Lucien Tesnière:

> The possibility of a term in the structural order having, beyond its unique higher connection, two or three lower connections [...] collides, in its place in a sentence, with the impossibility of a word in the spoken string being immediately in a sequence with more than two adjacent words

---

[14] The proposal of the knot model for syntactic representations was criticised by D. Krivochen (Krivochen 2018) who stays that there is no need to have a topology that allows for self-intersections in natural language descriptions.

> […] In other words, every structural node is susceptible to the creation of bifurcations, trifurcations, etc…, that are incompatible with linear order. […] There is thus a tension between the structural order, which has several dimensions […] and the linear order, which has one dimension. This tension is the squaring the circle of language. Its resolution is the sine qua non condition of speech. (Tesnière 1959, p. 21)

As Tesnière pointed out, "the impossibility of a word in the spoken string being immediately in a sequence with more than two adjacent words" proves the limitation of the linear phonic chain and can also be thought of as an impossibility to reduce the complexity of the linguistic meaning into linear strings. I will comment on this phenomenon in the next chapter.

## 5.3 NonReduction Theorem

It is a well-known fact that Peirce's philosophy and logic were inspired by his early studies of chemistry. At the age of only 11, Peirce attempted to write a History of Chemistry. He was summa cum laude in chemistry in 1863 at Lawrence Scientific School. It was his understanding of chemical compounds in terms of topology and relations, especially by creating diagrams or schemata of molecular relations, that Peirce applied later to his studies of logic (Ambrosio and Campbell 2017). In those years, chemistry was still acquiring its solid foundations and the status of a hard science. Works on topology by Peirce and other chemists in those years were special in that they tried to create abstractions that would enable them to classify similar chemicals together, which led to the later discovery of the Mendelian periodic table. The originality of Peirce's approach resided in that he observed the molecular relations rather than concrete chemical qualities; he was interested in molecular topology and its spatial representations.

Protein studies at present are in the same situation as chemistry in the second half of the nineteenth century; proteomics is only trying to establish its solid foundations. I propose to go back to Peirce's lesson in "chemistry of relations" (Ambrosio and Campbell 2017) and observe proteins with this optic. Chemistry provided a kind of material support for Peire's abstract logical reasoning. Now it may occur to proceed vice versa and use Peirce's abstract logical reasoning to approach proteins, biological material construing all forms of life. In what manner can Peircean theoretical thinking inspired by chemistry be applied to protein folding? As Robert Burch argued, "the most famous result in valental analysis is Peirce's Reduction Thesis." (Burch 1992)

As I illustrated, syntax intermediates in natural language between the linearity of sound and the non-linearity of meaning. Analogically, "protein grammar", or protein folding rules are an intermediation between the linearity of a peptide chain and the 3-D structure of a protein. I argue that by studying the linear sequences of amino acids, the deciphering of protein grammar is improbable. The reason why the study of sequence does not lead to comprehending the 3D structure (and experimental results promote this non-one-to-one correspondence) may be explained by the impossibility of reducing structure to sequence. In terms of Peirce's *Logic of Relatives*, a sequence is obtained by dyadic relations (bivalent elements are lined up in a sequence thanks to two blanks), but dyads cannot constitute anything more than a sequence.

> The extremely limited potential for relational construction starting from monads and/or dyads contrasts starkly with the immense, varied, and hugely interesting relational "world" available once triads are added to the initial stock of relations, as the following graphs indicate. (Burch 1992, p. 670)

Although the so-called NonReduction Theorem or Reduction Thesis was never explicitly formulated by Peirce, many Peircian scholars widely use this term (Burch 1992, 1997; Ketner and al 2011). The idea of (non)reduction was nevertheless clearly described by Peirce in many places in his work. In connection with the *Logic of Relatives*, a note taken from Peirce's essay *The Reader is Introduced to the Realatives* may be quoted:

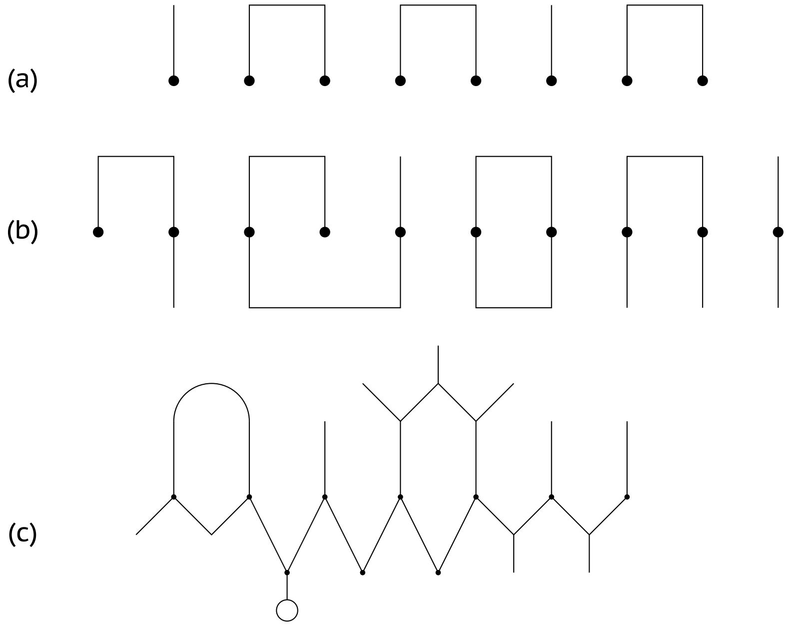

Figure 5.3: NonReduction Theorem. In the picture (a), structures are shown which can be obtained from monads (only dyads can be obtained from monads). In the picture (b) are structures which can be obtained from dyads (from dyads we can get a sequence). Yet we cannot obtain a more complex structure from dyads. Only from triads, such complex structures as in picture (c), are possible to obtain.

> Non-relative and dual rhemata only produce rhemata of the same kind, so long as the junctions are by twos; but junctions of triple rhemata (or junctions of dual rhemata by threes), will produce all higher orders. Thus, "– gives – to –" and "– takes – from –," give "– gives – to somebody who takes – from –," a quadruple rhema. This joined to another quadruple rhema, as "– sells – to – for –," gives the sextuple rhema "– gives – to somebody who takes – from somebody who sells – to – for –." Accordingly, all rhemata higher than the dual may be considered as belonging to one and the same order; and we may say that all rhemata are either singular, dual, or plural. (Peirce CP, § 3.421)

The idea of (non)reduction is also illustrated in additional enumerable Peirce's essays, with some authors (Brunning 1997, p. 252) mentioning his correspondence with Lady Welby, where the (non)reduction is expressed very clearly:

> I prove absolutely that all systems of more than three elements are reducible to compounds of triads... The point is that triads evidently cannot be so reduced. (Peirce and Welby 1977, p. 43)

Another quotation can be taken from CP Book III, Phenomenology:

> For were every element of the phaneron a monad or a dyad, without the relative of teridentity (which is, of course, a triad), it is evident that no triad could ever be built up. Now the relation of every sign to its object and interpretant is plainly a triad. A triad might be built up of pentads or of any higher perissad elements in many ways. But it can be proved – and really with extreme simplicity, though the statement of the general proof is confusing – that no element can have a higher valency than three. (Peirce CP, § 1.292)

Although one could continue with demonstrating Peirce's interest in the NonReduction Theorem and provide additional quotations, the demonstration of the validity of the Reduction Thesis is not the aim of this thesis and others already successfully did so, see e.g. (Burch 1992, 1997; Ketner and al 2011). The important point to make is that the idea of (non)reduction is closely related to Peirce's work in *Logic of Relatives* and is somewhat inherent and essential for all Peircian thinking, including metaphysics, semiotics and philosophy. The (non)reduction or, in other words, the teridentity of relation, is in the centre of all Peircian thinking. This is in all probability why Peirce himself never formulated it as a theorem, since it is inherently presented in his entire work. Nevertheless, the (non)reduction is also mentioned explicitly (see quotations above) and is exhaustively explained in the *Logic of Relatives* in terms of distinction between genuine and degenerate triads.

Adopting the antireductionist point of view proposed by Peirce, one cannot comprehend proteins in their complexity by reducing them to dyads (sequence). In the same manner, one cannot understand the verb “to give” as decomposed to dyadic relations between the donor and the receiver, the gift and the receiver and the gift and the donor. It is instead understood as a unique relation (teridentity relation), a simultaneous and integral act of giving. Protein as a 3D shape is reducible to elementary relations, but these are not of a dyadic character. A triadic understanding of relation in proteins can be understood as follows: peptide sequences are simultaneously folded and functional, exactly like a gift is given and received at the same time. A protein obtains its function by folding, it is folding because of its function and it functions because of its structure. The direction of dependency is being currently questioned in theoretical biology: Are protein folds Platonic forms (pre-given structures) or are they the results of natural selection of random folding (Denton, Marshall, and Legge 2003)? In other words, does the function determine structure or rather does the structure determine the function? This question is of course a chicken-or-egg problem and is meaningless for the purposes of this thesis. What is important to realize is that at this moment (evolution of proteins being completed for the current state of living beings) there is no random folding and proteins fold very quickly to obtain very specific and complicated structures. One could argue that folding and structure are interdependent and that the process of protein folding cannot be reduced to dyadic relations of peptide chain folding, structure creating, function obtaining: this action is simultaneous and integral.

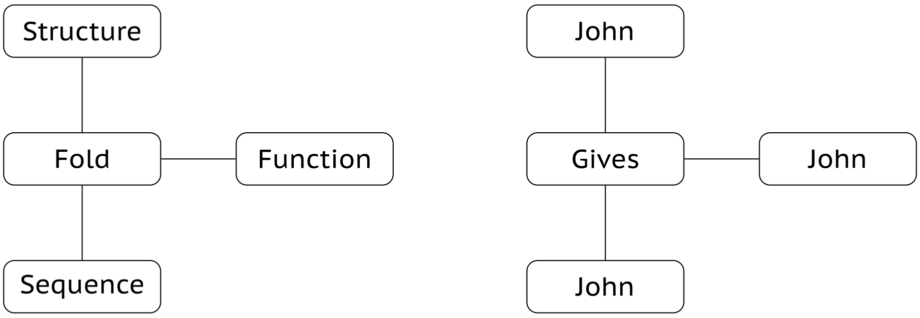

Figure 5.4: The teridentity relation. Analogy between the teridentity relation of the verb valency of the verb ”to give” and the teridentity relation of the protein folding.

Inspired by Peirce's lesson and understanding of triad as a primitive relative, it can be useful, in order to comprehend the "grammar of proteins" to find concrete primitive relatives in proteins, to find the crucial "empty unsaturated positions", "blanks" in the long peptide chain that are responsible for folding. The problem is that every amino acid is potentially triadic, considering the fact that they are macromolecules with many binding positions (peptide bonds and residue bonds), but not all amino acid residue have the blank that is used to create a meaningful fold. By meaningful fold, I mean the fold that determines the final protein shape and that is "a relation" in Peirce's terminology. Peirce distinguishes between "relation" and "connection": a relation is something more than a mere connection, since a connection is potentially between everything (also a non-connection is a kind of connection). Connection is therefore potentially between everything, is infinite (continuum). Relation, on the other hand, is not infinite but has a given number of unsaturated positions, a given number of relates. In a peptide chain, every amino acid residue is connected with others; this being the case of connection and not relation. Connection, in contrast with relation, has no informational or practical value, because if everything is connected, it gives us no information. This is why it is better to comprehend proteins as relations and not as connections (chains of connected amino acids).

> If relation were nothing but connexion of two things, all things would be connected. For certainly, if we say that A is unconnected with B, that non-connexion is a relation between A and B. Besides, it is evident that any two things whatever make a pair. Everything, then, is equally related to everything else, if mere connexion be all there is in relation. [...] This would reduce relation, considered as simple connexion between two things, to nothing. (Peirce CP, § 3.464)

Consider protein as a relative, that is, a relational structure with a given number of blanks (relates). Consequently, in proteins, the only binding places that give birth to a relation should be considered and not all binding places that are chemically presented in the entire peptide chain (since we are not doing a biochemical analysis, but a structural-topological one). The task is to therefore find elements that are crucial for folding, that are “relates”. The task is to find the limited number of binding positions in a protein. It seems improbable to find these binding positions that give birth to relations by merely observing a sequence of a peptide chain, at least from a relational (non)reduction point of view. In fact, alternative methods have already begun to appear which are seeking out the relationship between the structure and the sequence starting from the structure and not from the sequence, that is, that are not presupposing the reducibility of proteins to dyadic relations (Kister 2015).

# 6

# Participative opposition, content, expression

## 6.1 Asymmetric relations in language

In the previous chapter I briefly introduced the NonReduction Theorem by Peirce in order to argue for the limitedness of "the sequence to structure" approach in protein studies. Since the structure (3D shape) is not reducible to dyadic relations, the sequence as a chain composed of dyadic relations will never sufficiently express the 3D protein structure. One might argue, however, that the protein structure *is* reducible to dyadic relations. One cannot deny it because it is actually happening: the existence of the peptide chain being proof. My answer is that, as already mentioned previously, this is a case of *generating of* dyadic relations and not of *reduction.* In line with the theory of folding of continuum, I sustain that dyadic relations can be generated from a triad, yet cannot, in contrast, generate triads. It is true that in the process of protein synthesis, a protein is construed by the linear text written in amino acids language, which could be understood as a generation of triads from dyads. One should not, however, get confused. As I argued in previous chapters, the dyadic chain has to first be generated from a triad (since before it was only a continuous unarticulated line) and only after can it turn out to be a tool for creation of triads. This is still not a completely correct vision of the thing. It is true that the peptide chain serves as a starting point in the protein synthesis, but it is not the unique factor of creation of the final protein's shape. The context must come to finalize the entire process. We can admit that the dyadic chain generates somewhat the triadic protein, but only generates a *degenerated triad*,[15] which is to be completed by the contextual factors. *Generating* is not, however, the correct term. The peptide chain is a medium to express organic meanings, likewise language is a medium to express linguistic meanings. This does not mean that language generates meanings (although this can be pretended in generative grammar).

The very problem of the limitedness of a linear string in expressing non-linear meanings is easily demonstrable by ambiguous sentences (La vecchia porta la sbarra). Peirce was aware of this limitedness, which was why he preferred and proposed a complex tool of a non-linear way of expressing logical relations (Existential Graphs). Linear strings have, in contrast, an enormous advantage, having the power to encode and store for ages a huge amount of data with only a few letters (alphabets of natural language have around 26 letters, the amino acid alphabet has 20 letters). Linear strings are very economical and practical. The price for their economy is the fact that they are never exhaustive: an expression never exhaustively encompass the meaning it refers to. Of course, the exhaustiveness is even undesirable, at the moment when the expression exhaustively expresses the meaning, the two would equal, and thus the very role of expression would lose its sense. To be specific, take an example of a map as a representation

---

[15] "A Relation is either Genuine or Degenerate. A Degenerate Relation is a fact concerning a set of objects which consists merely in a partial aspect of the fact that each of the Relates has its Quality." (Peirce CP, § 2.91)

of a given territory. A map is a special type of representation, trying to match with the territory of the real terrain to the greatest possible extent. A perfect map does not exist, since a perfect map would have to represent *every single smallest point of the territory*, it would have a dipstick of 1:1, in which case it would lose its justification and cease to be a map.

Thus, the limitedness of expression is its desirable characteristics and the very nature of language lies in it. An expression cannot exhaustively encompass the meaning, but the meaning can exhaustively encompass its expression. Protein encompasses every one of the amino acids of the peptide chain and the bonds between amino acids do not break. It encompasses its expression, but has some additional value, having a biological function given by the shape and by the context. The territory encompasses every single point on the map, but also has some additional information. Apart from the points on the maps, it also has many more additional points. This is the fundamental relation between content and expression. Content encompasses its expression, it is a matter of a *participative relation* or *participative opposition.* This is the way C. Paolucci explains the relationship between content and expression: with aid of participative opposition as introduced by Hjelmslev (Paolucci 2010, p. 351). Over the next pages, I will try to explain the very notion of participative relation in language and consequently apply it to proteins.

## 6.2 The law of participation

Louis Hjelmslev borrowed the term "participation law" from the anthropology theory of Lévy-Bruhl and elaborated it in a more linguistic way as a very constitutive character of language. He elaborated this theory primarily in two essays: *La catégorie des cas* (Hjelmslev 1935) and *Structure générale des corrélations linguistiques* (Hjelmslev 1985). The core idea of the participation law governing a linguistic system resided in that, according to Hjelmslev, language is not analysable in terms of binary oppositions as opposed to the mainstream of structuralist linguistics of those years. Or, it is not analysable in terms of *exclusive* binary oppositions. Hjemslev himself, being one of the most influential structural linguists, did not deny binary relations in language, but did conclude, in a kind of compromise, that the binary oppositions governing language are not of an exclusive character, in other words, they do not exclude one another. To put it more simply, this means that the terms, in the participative opposition, may coexist without excluding one another.

Not all linguistic categories are definable, of course, in terms of participative oppositions. In phonology for instance, oppositions are always exclusive[16], one phoneme cannot be anterior and posterior at the same time, or labial and non-labial at the same time. The impossibility of superposing two contradictory features (distinctive features) at the same time guarantees the definition of a phonological unit: a phoneme is defined by exclusive oppositions, by phonemes with which it is in opposition. /p/ is /p/ because it is not /b/ with which it creates an exclusive binary opposition. /n/ is /n/ because it is not /ŋ/ with which it creates an exclusive binary opposition.

When we move, however, from phonology to "upper" linguistic areas, such as morphology or lexicon, the definition of linguistic units by means of exclusive oppositions becomes more complicated. It would be limiting to define, for instance, the word *man* by an exclusive opposition with the word *woman*, since the meaning of the word *man* is not limited to an opposition with the word *woman*. Imagine the following sentence: "All men are wise". In this case, the meaning of the word *men* is not definable by opposition to *women*. Hjelmslev observed, however, that in some way, there *is* a relation of opposition, and resolved it by calling this kind of opposition a *participative opposition* (borrowing the term from Lévy-Bruhl). The term *women* is in

---

[16] It should be remarked that, according to the phonological theory of the Prague school, the exclusiveness of the phonological oppositions is not that obvious. The Prague scholars suggested treating some of the phonological oppositions rather in a scalar manner than in terms of exclusive opposition. As a consequence, they differentiated between privative, gradual and equipollent oppositions in phonology (Trubetzkoy 1939, pp. 67–74).

| Case | Singular | Plural |
|---|---|---|
| nominative | žena | ženy |
| genitive | ženy | žen |
| dative | ženě | ženám |
| accusative | ženu | ženy |
| vocative | ženo! | ženy! |
| local | ženě | ženách |
| instrumental | ženou | ženami |

Table 6.1: Paradigm of the Czech noun žena. Syncretism in nom, acc, and voc in the plural forms.

opposition with the term *men*, but at the same time is *included within* the very term *men*.

This paradoxical situation is, according to Hjelmslev, by no means an exception within the language system and it is not merely a particularity of semantics or lexicon. In a like manner, all morphological categories are definable by participative oppositions. The case system, verbal tenses, gender and number of substantives, all these categories, enter into the participative oppositions. As a result, the accusative is opposed to the nominative, but is included within it at the same time. The plural is opposed to the singular but is also included within it. The past is opposed to the present but is at the same time included within it, etc. How is the accusative included in the nominative, the plural in the singular and the past in the present? Firstly, it is by syncretism of different functional units (nominative and accusative) in one morphological form (nominative) within a given paradigm. Secondly, it is by syncretism of different meanings (past, present) in one morphological form (present) in a specific context (the historical present for instance). In the case of the plural and the singular, one can also speak of metonymy.

The above-mentioned examples of a participative opposition between cases, tenses and the grammatical category of number illustrates the way language disobeys the laws of classical logic. By participative opposition, contradictory features coexist, with one and the same unit being accusative, vocative and nominative, all at the same time, without losing its identity.

> A linguistic system is free when compared to the logical system that corresponds to it. On the axis of the logical system it can be oriented differently, and the oppositions it creates is subject to the law of participation: there is not an opposition between A and non-A, the only oppositions in the linguistic system are between A on the one hand and A + non-A on the other hand.[17] (Hjelmslev 1935, p. 214)

Hjelmslev points out the anti-logical nature of the linguistic system, a system in which the basic prerequisites of Aristotelian logic (identity law, law of excluded third) do not work. Such a system is called a sublogical system. Hjelmslev claims that only a sublogical system can describe language phenomena. The core of the work of Hjelmslev lies in the opposition between an intensive (precise) term and an extensive (vague) term.

Hjelmslev is convinced of being a pioneer in the theory of participative opposition in linguistics. In fact, he claims that

> Our work is without precedence. Not only has our problem so far found no solution. It has not even been proposed. (Hjelmslev 1985, p. 30)[18]

Nevertheless, in the *Structure générale des catégories linguistiques*, he minutely analyzed previous works by the Russian linguists Peškovskij, Karcevskij and Jakobson, claiming that all of the three contributed to the problem of participative opposition. He was trying to prove their incompleteness: none of them went further in order to reach the very centre of the problem.

I am convinced that Hjelmslev elaborated the notion of participative opposition to an extent that no one had before, yet he is incorrect in

---

[17] In the original: "Le sytème linguistique est libre par rapport au système logique qui lui correspond. Il peut être orienté différement sur l'axe du système logique; et les oppositions qu'il contracte sont soumises à la loi de participation: il n'y a pas d'opposition entre A et non-A, il n'y a que des opposotions entre A d'un côté et A + non-A de l'autre."

[18] Original year of publication: 1933 in *Travaux du cercle linguistic de Copenhague.*

attributing originality to his work, since he himself in other places, perhaps due to distraction, equates his own terms with the terms by Peškovskij. He demonstrates with the example of the nominative, that the notion *catégorie zéro* by Peškovskij expresses the same meaning as his own notion *term extensif*.

> The nominative is to the cases what the present is to the tenses. The nominative is what Peškovskij called catégorie zéro. (Hjelmslev 1935, p. 100)[19]

The difference between *catégorie zéro* and *extensive term* is, according to Hjelmslev, in the fact that Peškovskij states that *catégorie zéro* has no signification, in other words, it is characterized by the lack of meaning. An intensive term, on the contrary, has a signification which is guaranteed by subsumption to a category: to belong to a category assures already a certain meaning. Does Peškovskij actually claim, however, that *zero category* has no meaning? Isn't it merely a rhetorical figure to argue that the meaning is extremely vague? Peškovskis states: "lack of meaning constitutes a meaning in and of itself".[20]

## 6.3 Jakobson and markedness

In continuation with Peškovskij, Jakobson elaborated the idea of lack of meaning being meaningful in his essay *Signé Zéro* (Jakobson 1971b). He replaced the notion of *catégorie zero* with a more general concept of the *signe zéro* (zero sign). According to Jakobson, one can talk about two basic types of *signe zéro*, where the former concerns the signifier and the latter concerns the signified. He argues that the so-called *signe zéro* represents one of the elementary morphological traits of a linguistic system: nominal and verbal systems are decomposed into binary relations between a term indicating a mark (a marked term) and a term that does not express either the presence or the absence of the mark

---

[19] In the original: "Le nominative est aux cas ce qu'est l'indicatif aux modes, le présent aux temps. Le nominative est ce que M. Peškovskij a appelé une catégorie zéro".

[20] cited from (Jakobson 1932, p. 3)

(*signe zéro* or unmarked term). This case is different from the classical concept of the *signe zéro* in phonology, where the *signe zéro* signifies the absence of a feature, for instance sonancy. A voiceless phoneme is in a relation of binary opposition with a voiced phoneme. The kind of *signe zéro* that concerns the signified (Jakobson 1971b, p. 212) is a case, however, where the opposition is not between an absence or presence of a certain feature (mark), but between the presence of a mark and non-information about its absence or its presence. This is a case of opposition between, for example, the perfective and imperfective verbal aspect or between the grammatical gender.

Jakobson's essay is very confusing and not intuitive to understand, since he treats both kinds of *signe zéro* in the same manner. The fact is that the difference between the two is of great importance within grammar theory and it is unfortunate that Jakobson did not expand on the difference between the two types of *signe zéro*. Very briefly, the basic difference resides in that the first one regards the very linguistic form (phonological or morphological) while the second one regards the meaning. The choice of terminology is confusing as well: it is confusing to apply the term "zéro" to the level of content, since it somehow contradicts the very nature of the *zero sign* concept. A *zero sign* concept states that even if there is no signifier, it actually means something. Even a zero may have a meaning (thanks to a paradigmatic opposition with other signifiers). Thus, if Jakobson argues that zero is the meaning of a grammatical category (imperfective, masculine gender, etc), it is in order to argue for the contrary of the original zero sign concept: it says that there are signifiers that have no meaning. What would, however, such a signifier be good for? Why would anyone want to express *nothing*? The problem is in the choice of terminology which is certainly influenced by Peškovskij's term *catégorie zéro*. For Peškovskij, however, it was not the case that catégorie zéro had no meaning, but on the contrary, he argued: "lack of meaning constitutes a meaning in and of itself".

Let us return, however, to Hjelmslev. Apart from the ambiguity of the meaning of *catégorie zero*, the difference between Hjelmslev and the Russians is quite blurred, although Hjelmslev was convinced there was an important novelty in his theory. After a minute analysis of the above-mentioned works, it is true that both the Russian theories and Hjelsmlevian participative opposition seems quite alike. The law of participation (loi de participation) finds an equivalent in Russian theories in the law of substitution (loi de suppléance). Jakobson uses the term *concept of transposition* (Jakobson 1932, p. 2). The law of substitution, or *concept of transposition*, is explained by Jakobson with aid of the following example.

> The Russian word *osël* "donkey" contains no indication of the sex of the animal in question. If I say *osël* I make no decision as to whether I have to do with a male or with a female, but if I am asked *èto oslíca*? 'Is it a she-ass?' And I answer *nét, osël* "no, donkey", then in this case the masculine gender is indicated. (Jakobson 1932, p. 2)

The law of substitution resides in that a grammatical category might be substituted by another grammatical category. In such a manner, the masculine gender substitutes the feminine gender: the Russian feminine substantive *oslica* may be substituted by the masculine substantive *osël*. Likewise, the English word "man" can substitute the word "woman", for instance in the sentence "All men are wise". As concerns grammatical categories, the situation is analogical. Speaking of verbal tenses, the present (the so-called historical present) substitutes the past, for example in the Czech sentence

(16) Dobrá tedy, teď si trochu zacestujeme v čase. Je rok 1970, lidé používají dvě lampičky, jedno rádio, troubu a pevnou linku.[21]
'Well, now, let's travel in time a bit. It is the year 1970, people use two lamps, one radio, oven and a fixed line.'

or the future, for example in the Italian sentence

(17) Domani vado a scuola.[22]
'Tomorrow I go to school.'

---

[21] The sentence is taken from the Czech National Corpus SYN 2015 (Křen et al. 2015)

[22] The sentence is taken from corpus of Italian language BADIP (Bellini and Schneider 2003–2018).

Speaking of the category of number, the singular number in many cases functions as a substitution for the plural number, speaking of the category of case, the nominative functions as a substitution for the accusative for instance, etc. Many other examples may be found amongst linguistic categories. In order, however, for a grammatical category to be substituted by another one, there must be a relation of participation between the two. *Osël* and *oslica* are in a relation of participation. The law of substitution in fact requires the law of participation, or said otherwise, the participation law is a prerequisite for the substitution law. A linguistic category may replace another linguistic category only if the latter category participates in the former. A linguistic category (a vague term) may substitute, in specific contexts, another category (precise term), because the vague term, among all the meanings it encompasses, may also express the meaning of the precise term. So to say Hjelmslev's notion of law of participation is more about the very nature of language, while the Russians' notion of law of substitution is more about the practical consequences of the law of participation: as a principle of economy, the law of substitution permits the substitution of the accusative by the nominative, substitution of the past by the present, or the substitution of the feminine by the masculine.

Another point which Hjelmslev believes to differ from Russians is the possibility of applying his theory on systems of more than two elements, while Russians only applied it to a system of two elements: vague/precise, marked/unmarked, etc. Hjelmslev demonstrated how the theory of participative opposition might be applied to a system even as complex as the case system (Hjelmslev 1935). There is, however, a dualism at the heart of his method and a dualistic opposition between the extensive and intensive term.

The truth is that Hjelmslev in his formula of the participative opposition mentions three terms: A VS A plus non-A, while Russians only use two terms: A VS non-A. This amounts to nothing, however, but an illusion and the problem is actually conceptual rather than real. In fact, Jakobson with the unmarked term (non-A) meant all other things that are a contradiction or contrary of A. Jakobson inspired his mor-

| | | | Variantes of Term II | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Term I | Term II | a | b | c | ab | ac | bc | abc |
| a | + | + | + | - | - | + | + | - | + |
| c | - | + | - | - | + | - | + | + | + |
| b | - | + | - | + | - | + | - | + | + |

Table 6.2: Hjelmslev's participative opposition .The participative (dualistic) opposition between the extensive and intensive and theoretic variants of the extensive term (Hjelmslev 1935, p. 40).

phological work on Trubetzkoy's system of phonological oppositions (Trubetzkoy 1939), there is a significant difference between the two systems however and Jakobson demonstrates it in his essay *On the Structure of the Russian Verb* (Jakobson 1932). Jakobson applies the phonological structure to the morphological one, although, as of the very first page of his essay, Jakobson calls attention to the fact that the opposition in morphology has a specific nature, which allows the existence of a term that has an internal contradiction. This means that this term, the unmarked term, may have different and even contradictory meanings. The contradictory nature of the unmarked term is resolved thanks to context: even if the meaning in itself is ambiguous or contradictory, there is no ambiguity once the context is given. Jakobson provides an example of the Russian substantive *osël* (donkey). *Osël* is grammatically a masculine form, but it can also have a general meaning. Thus, depending on the context, the word *osël* may mean "donkey-masculine" or "donkey-feminine" or "both genders of donkey" or "it is of no importance whether feminine or masculine". The feminine substantive *oslica*, on the other hand, only has one single meaning, which remains unvaried in every context, with this meaning being "donkey-feminine". *Oslice* is a marked term. When we therefore place the Russian words *osël* and *oslice* in the formula of a participative opposition, we do not end up with a contradictory relation, but *a relation of participation.* As a consequence of this relation, the word *osël* may substitute the feminine form *oslica* (law of substitution).

The Czech linguist Miloš Dokulil (Miloš 1957) represented Jakobson's formula as illustrated in the picture:

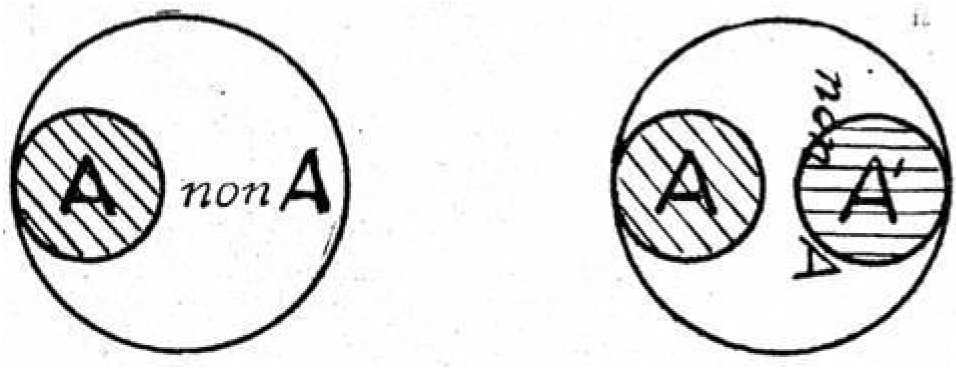

Figure 6.1: Jakobson's opposition between a marked and unmarked term.A is a marked term, A' is a term contrary to A.Non-A is the unmarked term that occupies the entire space. Picture by Miloš Dokulil. http://as.ujc.cas.cz/archiv.php?art=936

As is clear from the graph, the space that "non-A" occupies, is the entire space that also encompasses A and also A'. Thus, a participative relation is illustrated between A and non-A.

The difference between phonology and morphology, even both using the same formula A VS non-A (marked VS unmarked) is fatal. There is no space for the law of substitution in phonology. The lack of a mark is meant as a missing element, which by its non-presence distinguishes between two phonemes. In morphology, in contrast, no lack of a mark is mentioned, there being only an indecision as to whether the mark is or is not present. To put it differently, *there is no information about the fact whether the mark is or is not expressed.* The unmarked term is *undecided.* Jakobson states it clearly: "there is an antinomy between the non-signalization of A and the signalization of non-A." (Jakobson 1958/1984, 13).

Jakobson also distinguishes between two types of unmarked terms in morphology: the first one is the case of participative opposition, the second one is the case of exclusive opposition. The Russian word *osël* might happen to be in both types of unmarked terms: the second one (exclusive opposition to the marked term) concerns a particular case which has to be determined by context. The particularity of Jakobson's approach resides in that he considers the context as making up part of his grammatical theory. This is a special characteristic of the Prague school, which Jakobson was part of. Unlike other schools of structuralism, the Prague school paid a great deal of attention to the role of communication, actual language use (parole) and not only language in a systematic way[23]. Language as a dynamic and unstable system, even within synchronic studies, was proposed by V. Mathesius, the founder of the Prague school.

To conclude this chapter, the point that Jakobson made with his double nature of an unmarked term (restricted and general sense) leads us in a direction already intimated by the theory of the folded continuum: a term undecided (contrary or contradictory) on the level of system (langue) may become decided, that means dyadic on the level of language in actual use (parole). *Osël* may have a restricted sense of "donkey masculine" only in a specific context which is created by a "specific fold" of the linguistic system. Thus, the contradictory nature of a linguistic system becomes unambiguous at the moment of speech, of the use of language. The moment of speech is a concrete fold of the continuum. Only when continuum is folded, a particularity, an identity arises. At the moment of the folding, ambiguity disappears, *osël* is donkey masculine or it is both masculine and feminine, but it is one of the two possibilities, not two at the same time.

In summary, a seemingly desperate situation of the ambiguity of the linguistic system outlined by the participation law is elegantly resolved when one moves from the system (langue) to the linguistic use (parole). We are slightly touching at this point the field of pragmatics. In the following section, I would like to extend the theory of participation, as presented by Hjelmslev, to the larger semiotic context and consequently to the context of protein folding. I would like to demonstrate that it is closely connected to the theory of folded continuum in proteins.

---

[23] I can cite the celebrated work by Skalička entitled *The Need for a Linguistics of "la parole"* (Skalička 1948).

## 6.4 Content and expression

The aforementioned examples of extensive terms (*osël,* present tense, nominative, singular of substantives, etc) were intended to describe the constitutional nature of the content set, whether at a meta level (the meaning of morphological categories) or within a single system (single words). The nature of the content set is similar to the expression set, and this nature is made by the binary oppositions. In contrast to the expression, the oppositions for the content are not necessarily exclusive, being between positive and negative terms (A vs B), or between marked and unmarked[24] terms (A vs non-A). Hjelmslev places intensive term and extensive term in opposition, the former being the concentrated meaning of the latter. Opposition is therefore participative and not exclusive and can be expressed by the formula "A vs A + non-A", where "A" is the intensive term and "A + non-A" is the extensive term.

Thus far we have seen the nature of the content set characterized by participative opposition. The idea of participative opposition by Hjelmslev can be used, however, as a useful tool not only to define, by means of opposition, the content set, but can also be used to define the very nature of the relationship between the content and the expression. As was mentioned at various times in this thesis, the relationship between the content (meaning) and the expression (sound) is extremely hard to approach and understand. This relationship primarily concerns asymmetry between linearity and non-linearity. In the words of S. Karcevskij, we could use the term "asymmetric dualism between the signified and the signifier" (Karcevskij 1956). Participative opposition, or the law of participation, is one of the possible means of resolving the problematic definition of the relation between content and expression. C. Paolucci already proposed extending Hjelmslev's

---

[24] In this case, I use the notions of marked and unmarked term as introduced by Trubetzkoy (Trubetzkoy 1939) in phonology, which means terms constituting exclusive opposition. For Jakobson's studies on the morphology of the Russian verb, in contrast, the terms marked and unmarked were used as constituting a kind of participative opposition (similar the case of osël and oslice).

formula of the participative opposition to the correlation between expression and content. According to Paolucci (Paolucci 2010), expression and content are in such a relationship, as intensive and extensive terms. The law of participation explains the problematic question of *how the passage from a linear string (sound) becomes a complex structure impossible to be expressed linearly (meaning).* The meaning of a sentence, for example, is a complex structure impossible to define by means of linearity. Even if the sentence is expressed by a linear phonetic chain, to understand the complexity of the meaning of the sentence, all relations between the single constituents, a non-linear syntactic representation is needed. One can therefore argue that the phonetic chain is encompassed within the meaning of the sentence, it *participates* in the meaning. The phonetic chain, obviously, participates in the meaning, but is not sufficient to decipher the entire meaning, other factors having to be added such as syntactic relations and pragmatic factors. To put it in a different way, expression is a condensed version of the meaning, it is the *intensive* term. This approach allows us to comprehend the semiotic notions of content and expression not as two separate entities, but as being *part of* each other, undissociable, as the celebrated metaphor of de Saussure states, two sides of one sheet of paper.

> Thought is the front and the sound the back; one cannot cut the front without cutting the back at the same time. (de Saussure 2011, p. 113)

Yet the law of participation goes beyond the metaphor by de Saussure. The metaphor is, as a matter of fact, misleading, since it implies that content and expression are equal entities in the sense that both represent sides of a paper sheet, which are of the same size and material. Paolucci's idea of applying the participative opposition to the relationship between the expression and the content expresses, however, the asymmetry between the two, inequality, content and expression are not of the same rank, are not of the same degree. The participative opposition between content and expression consequently expresses reciprocal undisociability on the one hand, and asymmetrical inequality on the other.

This short preview of Hjelmslev's theory of intensive and extensive terms and its extension to the semiotic function itself between content and expression leads us to consider the possibilities of differential and relational definitions, not only in one of the semiotic planes (expression), but also in the content plane and, lastly, in the character of the function itself between expression and content. By differential and relational definitions, I mean the definitions in terms of binary oppositions. The participative opposition, being also a kind of binary opposition, creates a compromise between strict binary structuralist thinking and semiotic theories that deny the binary character of semiotic phenomena (Peirce).

Paolucci's understanding of expression and content provides a strategic starting point for reconsideration of classical semiotic theories (De Saussure for instance).

## 6.5 Consubstantiality and proteins

In the previous section I introduced the application of participative opposition to the semiotic relation between content and expression. It was intimated that this approach helps reach an understanding of the problematic relation between content and expression, between linearity and non-linearity, this being problem I was dealing with since the first pages of this thesis. The problem of passing from the linearity to non-linearity is analogical in linguistic studies and in protein studies. In the protein folding process, the linearity of the peptide chain becomes the dimensionality of the functional protein. The linear string of the peptide chain does not disappear, however, but continues to exist, in the form of a folded linear string. In similar fashion as the linear phonetic chain makes up part of the sentence meaning, the linear peptide chain makes up part of the protein. The linear peptide chain is encompassed by the protein, is part of it and *participates* in its identity.

One could argue that there is a relation of participation between the protein and its peptide chain not only because the protein subsumes the peptide chain, but also because proteins and their peptide chains are substantially identical. They share one mutual substantiality (amino acid string). In this manner, the participation between the organic content and its expression is governed by the law of consubstantiality. Consubstantiality is a term introduced by Lévy-Bruhl, who is the author of the very concept of participation in language (and in culture in general). Lévy-Bruhl assigned the feature of participation to language because he was convinced that language makes up part of culture and ultimately that culture is connected to the mind. In this manner, he spoke of the "primitive mind" in primitive societies. He explained the difficult attitude of primitive people towards their own images by the very notion of consubstantiality and participation:

> When a savage sees his own image (shadow, reflection, etc.) it is not a more or less faithful reproduction of his features, it is the consubstantiality that he imagines and feels between them and him. But he can also imagine and feel this essential participation between him and a being whose external appearance is different form his own. (**lévy1996**)[25]

The consubstantiality is not only a concern of the visual perception of images in primitive societies, it is a phenomenon that goes far beyond perception. It goes to the very mental processes and is also reflected in the language of primitive societies. As an example, the "primitive" denomination of things in the world even becomes incomprehensible to our understanding. In the culture of Australian aborigine people, for example, the name for the sun and the name for a white cockatoo are considered to have one shared meaning. This is not a simply synonymy, however, the fact is that the very signification of the word is the sun and the white cockatoo at the same time. The concepts of the sun and the white cockatoo are, in this culture, consubstantial, in other words, for the aborigine people, a situation where something is

---

[25] In the original: "Le primitif quand il voit sa propre image (ombre, reflet, etc.) ce n'est pas la reproduction plus ou moins fidèle de ses traits, c'est la consubstantialité qu'il imagine et q'il sent entre elle et lui. Or, cette participation essentielles, il peut aussi l'imaginer et la sentire entre lui et un être dont l'apparence extérieure est autre que la sienne."

the sun and at the same time the white cockatoo, is absolutely normal. For aborigine people, the terms sun and white cockatoo do not exclude each other (Harris 2009).

Although the research of Lévy-Bruhl was primarily dedicated, of course, to anthropological studies and not linguistics and if he did study linguistic systems, it was linguistic systems of primitive people, Hjelmslev succeeded in applying the notion of participation to the linguistic system in general. Hjelmslev considered not only primitive languages, but also modern European languages to be governed by the law of participation whose main trait is that it does not correspond to the laws of classical logic. As a matter of fact, language makes up part of culture and there is a continuity between the language of primitive people and modern languages. As a consequence, all modern languages carry the heritage of primitive languages and as a consequence, all grammatical systems are far from being logic.

> Moreover, M. Lévy-Bruhl admits that the difference between the primitive mentality and ours is far from absolute. Every human mentality is marked by a certain primitiveness. The representations and the connections of the representations governed by the law of participation are far from having disappeared. (Hjelmslev 1928, p. 262)[26]

The law of participation, characterized by consubstantiality, disobeys the classical logic law of the excluded middle. As a consequence, the sun is at the same time the white cockatoo, the future is at the same time the present and the nominative is at the same time accusative. There is nothing inconsistent, however, in this observation, modern languages have simply inherited certain features from primitive cultures; this is how Hjelmslev explains the illogical nature of languages. In proteins, consubstantiality is more evident in respect to natural language, with the linear peptide chain clearly sharing its substance with

---

[26] In the original "D'ailleur, M. Lévy-Bruhl admet que la différence entre la mentalité primitive et la nôtre est loin d'être absolue. Toute mentalité humaine est empreinte d'une certaine primitivité. Les représentations et les liaisons des représentations regies par la loi de participation sont loin d'avoir disparu."

a functional protein. If the two parts of the semiotic function (expression and content, or, peptide chain and protein) share, however, the substance thanks to the law of participation, what is the formal difference between the two? The only possible answer is that the formal difference, guaranteed by the double articulation, is the concept introduced by the French linguist A. Martinet (Martinet 1967) and is extended to semiotics by Deleuze and Guattari. The double articulation actually treats one and the same substance as disposed to be articulated (*folded*) in two steps, where the first step affects content (articulation of phonemes, which, by combinatory rules, create words and sentences) and the second step affects expression (articulation of single phonemes). In the organic context, the first articulation affects the folding of the peptide chain and the second articulation affects the individuation of single amino acids. The distinction between the two articulations is not, as pointed out by Deleuze, between forms and substances, but between content and expression:

> A stratum always has a dimension of the expressible or of expression serving as the basis for a relative invariance; for example, nucleic sequences are inseparable from a relatively invariant expression by means of which they determine the compounds, organs and functions of the organism. To express is always to sing the glory of God. [...]The first articulation concerns content, the second expression. The distinction between the two articulations is not between forms and substances but between content and expression, expression having just as much substance as content and content just as much form as expression. (Deleuze and Guattari 1987, p. 43)

Deleuze's understanding of the terms *content* and *expression* is traditional in the Hjelmslevian sense in that he comprehends the two sets of semiotic relation as being interdependent, in other words, only being definable by reciprocal solidarity and in no other way (having no identity as single entities, but only in relation to each other). Deleuze's understanding of the terms *content* and *expression* is untraditional, however, in that the two sets, in this understanding, do not have separate forms and separate substances. For Hjelmslev, in fact, four separate units were defined (Hjelmslev 1963):

- form of expression
- substance of expression
- form of content
- substance of content.

According to Deleuze, however, *expression has just as much substance as content and content has just as much form as expression.* This means practically that there is no such difference between content and expression that regards the form and the substance. Or, at least, there is no *quantitative* difference between the two sets, inasmuch as *expression has just as much substance as content and content has just as much form as expression.* One might then ask, however what the difference is between content and expression? According to Deleuze, this difference is only and exclusively in the articulation. The difference is in the way of articulating the formed substance. By articulation, in the case of protein folding, I mean the singular folds, the folding of continuum. The difference guaranteed by articulation is important since the first articulation is much more complex than the second articulation.

> There is never a correspondence or conformity between content and expression, only isomorphism with reciprocal presupposition. The distinction between content and expression is always real, in various ways, but it cannot be said that the terms preexist in their double articulation. It is the double articulation that distributes them according to the line it draws in each stratum; it is what constitutes their real distinction. (Deleuze and Guattari 1987, p. 44)

The "line" Deleuze refers to, is the "cutting line" that cuts the continuum, in our case it would be more opportune to talk about "a folding line", as I proposed previously, inasmuch as, in our case, the transition between organic strata is given due to the fold of the peptide chain. Deleuze's observation that "it cannot be said that the terms preexist their double articulation" and that "it is the double articulation that distributes them according to the line it draws in each stratum" leading us back to the argumentation of the impossible preexistence of the expression without a connection to the content, already discussed in the second chapter. It is thanks to the double articulation that the consubstantial linear chain appears both at the level of content and at the level of expression. The analogy between the phonic chain and the peptide chain is very clear in this case: the first articulation, protein folding, affects the content and the second articulation, the individuation of single amino acids, affects the expression.

At this point it is worth mentioning the work of S. Ji, who elaborated in detail the application of the concept of double articulation to protein folding. The first articulation is also attributed to the protein folding in papers by Ji and the second articulation is attributed to single amino acids forming the peptide chain:

> Also called double articulation, duality is considered one of the most fundamental characteristics of all languages [...] The first articulation is responsible for the almost infinite number of sentences that can be generated from a finite number of words, obeying a finite set of combinatorial rules called grammar. The second articulation refers to the formation of words by combining simpler units, phonemes [...] and morphemes [...]. Cell language also possesses duality, and I attribute this ultimately to the duality of molecular interactions, namely, covalent (also called configurational) and noncovalent (or conformational) interactions. Examples of covalent interactions include linking nucleotides or amino acids to form nucleic acids or polypeptides, respectively, and phosphorylation and dephosphorylation reactions in signal transduction pathways. Examples of conformational interactions are provided by folding reactions of biopolymers and binding reactions between ligands and receptors and between transcription factors and DNA regulatory sites. (Ji 1999, p. 24)

It is apparent that Ji's usage of the notion of double articulation is absolutely coherent with the linguistic tradition, the first articulation being attributed to larger languages units as sentences and the second articulation attributed to the smaller language units, phonemes (or morphemes). Analogically, applied to proteins, the first articulation being attributed to larger units, proteins, and the second articulation being attributed to smaller units, amino acids.

With the aid of Deleuze's treatment of the double articulation as accordingly concerning the content (first articulation) and the expression (second articulation), we can conclude that in the case of protein folding, it is semiotically convenient to approach functional proteins as contents and their peptide chains of amino acids as expressions. I place the stress on the specification of the case (protein folding), because in another case the situation might be different: semiotic units are not frozen entities, but always definable and redefinable in accordance with relations and context. Moreover, the discussed notion of double articulation can be connected to the theory of folded continuum: the first articulation is represented by the folding of linear continuum (peptide chain) while the second articulation is represented by the binding together of basic units constituting the very linear chain (amino acids). Note that also in the case of Martinet' s double articulation, similarly to the Peircian theory of continuum, the direction goes from the non-linearity to linearity, in other words, from triadicity (dimensionality) to dyadicity (flatness). Firstly, the more complex units are defined and only consequently basic units emerge. It is important to realize that there is a fundamental contrast between the approach described above, on the one hand and the generative approach represented by classical generative syntax, a contrast I commented on in detail in the previous chapter.

In summary, my aim in the prior chapters was to revise some of the classic linguistic theories, such as for instance syntactic representations or the theory of double articulation, in light of the theory of folded continuum. My viewpoint is a semiotic one and results from a presupposition that all grammatical description has to deal with one and the same problem in the end, this being the issue of the tension between linearity and non-linearity, represented by the two sets of the semiotic function (expression and content). The solution I propose concerning how to treat grammatical phenomena is to

1. always consider the two sets of semiotic function as interdependent, which means, do not approach only one of them independently from the other,
2. begin the description from the non-linearity (content) and proceed towards the linearity (expression), while comprehending the expression as the consequence of the folded content.

# 7

# From content to expression

## 7.1 Meanings in the biosphere?

We have seen in the previous chapter how the notions of content and expression might be connected to the notion of double articulation. The double articulation was assigned not only to natural language, but also to protein synthesis (following the examples of Deleuze and Ji), which imply that the analogy between natural language and protein synthesis, which has been at the centre of this thesis from the first pages, moves in the direction of treating functional proteins as contents and strings of amino acids as expressions.

My aim was not, however, to state that content in the sense of linguistic content (meaning) exists at the level of proteins or cells. My aim was to instead demonstrate that when dealing with natural language on the one hand and protein synthesis on the other hand, it can be argued that the constituting principle of both is *a code* (in terms of

| Articulation | Natural language | Protein synthesis |
|---|---|---|
| First (content) | Words and syntagms | Secondary structures, domains, proteins |
| Second (expression) | Phonemes | Amino acids |

Table 7.1: The two articulations in natural language and in protein sythesis

the semiotic definition mentioned in Chapter 4) and that there exists an isomorphism in the structural design of the two codes. This design resides in the decomposability of more complex and non-linear units into less complex and linear units. The isomorphism does not mean that protein structures *mean something*. The isomorphism is not between proteins and meanings, but the isomorphism is between the non-linearity of folded protein and the non-linearity of linguistic meanings.

One cannot avoid, however, facing of the question of whether there is the equivalent of a linguistic *content* in the living systems? Can we speak about *meaning* at the cellular level?

## 7.2 The lower threshold of semiotics

The problem of attribution of the term *meaning* to living cells and organisms requires a kind of consciousness, or a problem of attribution of agency. It is, in other words, intentional behaviour, which is not only a point of interest for semioticians or biosemioticians, but has also been widely discussed among biologists. I might mention the celebrated Nobel Prize talk by Barbara McClintock from 1983 who outlined cell consciousness as a challenge for future research.

> There must be numerous homeostatic adjustments required of cells. The sensing devices and the signals that initiate these adjustments are beyond our present ability to fathom. A goal for the future would be to determine

> the extent of knowledge the cell has of itself, and how it utilizes this knowledge in a “thoughtful” manner when challenged. (McClintock 1984)

U. Eco led a discussion on a similar topic with Italian immunologists, with one of them, a discussion with Italian biologist G. Prodi, becoming particularly well-known. Eco disagreed with the attribution of semiosis to lower forms of life, but finally immunologists persuaded him to accepting that one can speak, on a cellular level, of a kind of elementary semiosis. Eco placed elementary semiosis in the *lower threshold of semiotics* (Eco 1990, p. 228). Eco doubted, however, that cells could have interpretive options, refusing to assign the Space C (Space of interpretation options) to them, and arguing that the way cells communicate and interact with the environment, works according to the stimulus-response scheme. In his later works, especially in *Kant and the Platypus* (Eco 1997), Eco classified such processes as *primary icon*s. This concept is not unique to cell processes, however, as it has even been assigned to humans, in particular to human perception (Eco 1997, 2007).

Primitive semiosis exists both in primitive forms of life, as well as in human organisms. Our body performs automatic activities every day that work according to the stimulus-response scheme. The activities that we do instinctively, unknowingly, do not only concern life functions such as breathing, processing and transmitting nutrients and the like. Instinctively, we also perform everyday routine sensory-motor activities, avoiding automatically, for example, obstacles and walking people on our walkway. We also instinctively perform cognitive activities such as recognizing the way home from work: moving the familiar pathway, we go, so to say, blindly, not paying attention to when to turn and when to cross the street, often not being aware of how we found our way home.

Apart from the aforementioned automatic activities, people daily perform complex interpretive activities: when we have to deal with problems of various kinds and when Space C opens before us. The question is whether Space C exists in primitive forms of life and whether these have an option of “free” interpretation and the possibility of decision-making. I have already mentioned that U. Eco was sceptical about this claim. Eco admitted, however, after a long discussion with immunologists, that there is a certain space for contextual decision-making for immune cells, for example, when a lymphocyte meets an antigen. Although this space is incomparably smaller than Space C for humans, it is still there. (Eco 1990, p. 222)

In the biosemiotic paradigm, it is believed that there is such a semiosis in cells (and in every organism in general). K. Kull, for example, even claims (Kull 2015) that evolutionarily, before the organisms reach a stage where they instinctively “know” which of the several options to choose, in other words before the organisms acquire instinctive behaviour (and when Space C will disappear, only one option of behavior remains) they make kinds of decisions or choices and, on the basis of experience, determine whether the choice was successful and, if it was, they repeat the same choice in each of the following analogical situations, until it finally becomes automatic and sometimes partly inheritable. Kull calls such a process “learning” or “semiotic scaffolding”, a sort of semiosis, which can subsequently become a “fixed” code, where Space C temporarily “disappear”.

The oscillation at the interface between the strict “fixed” code and the free Space C of interpretation makes up part of our everyday life: we automatically continue driving when the light is green. According to the code rules, we simply continue driving without a conscious evaluation of the situation, our action being automatic. When there is a situation when the light is green, and a pair of schoolchildren are crossing the street at the same time, we deflect the automatic behaviour: space C has been opened, offering at least two possibilities, to obey the traffic rules or not to obey the traffic rules, or to put it in a different manner, to obey common sense and stop or obey the traffic rules and not stop. There could be more possibilities of course, more than two, for example to honk and stop, to stop, to open the car window and reprove children, etc.

Do also the lower forms of life deal with situations that have not foreseen code rules? Does Space C open up in front of them? While it is undeniable that animals and inferior forms of life act instinctively, they also get into situations where they can (and often are forced in order to survive) to disobey the rules of the code they were following, or to introduce a new code. Bacteria resistance to antibiotics, for example, is generally speeding up, which means it has been able to develop an alternative response when meeting with a given substance.

Space C, or in the words of K. Kull, "the phenomenal present" (Kull 2015), appears in the biosphere. The coexistence of multiple options, faced by a living entity, is a prerequisite for what S. Kauffman (Kauffman 2000) described as the impossibility of determining the biosphere configuration space. According to Kauffman, cells are constantly expanding the space of possible actions in the near future, but these possible steps depend on "the phenomenal present" which provides more than one option, since it is impossible to estimate which further consequences will arise.

In conclusion, to face the question as to whether there are meanings (in the sense of linguistic meanings) in the biosphere, would require defining what linguistic meanings actually are, in other words, what can be defined as the study matter of semantics. One could actually argue that there are so many definitions of meaning (propositional meaning, intensional meaning, extensional meaning, rhetorical meaning, etc) that which of the meaning definitions I am working with needs to be specified. The answer to this question is not my focus of interest, however, in this thesis. I can only state that, in order to deal with meanings, whichever definition of meaning one chooses, there are at least two conditions:

1. There must be Space C, no matter of what dimension

2. There must be the possibility of choice-making, that is the possibility of "resolving" Space C

In my view, meanings are dynamic variables that require interpretative options. In a similar manner, meanings are also approached by Kauffman (Kauffman 2000, pp. 109–110).

## 7.3 Chance, (Love) and Habit

I introduced the notion of the phenomenal present and Space C in order to comment on the possible interpretation in living systems and as an attempt to answer the question about meanings in the biosphere. The existence of Space C is guaranteed by an incompossibility, the coexistence of several options at one and the same time (the phenomenal present). The organism, at the moment of interpretation, faces this incompossibility of options and has to choose only one of all possible options. Of more interest, however, for the purposes of this thesis, is not the capacity itself of making interpretations, but the capacity of making *codified* interpretations. Put more simply, an interpretative capacity itself is not the only condition for a semiosis, for a code-based behaviour. Oscillating between the freedom of interpretation and frozen rules of acting is the key feature of semiosis.

To approach more deeply the oscillating between the two extremes, it is worth mentioning the work of Ch. S. Peirce (Peirce 1956). Peirce's notion of *chance* corresponds to the choice of one of several incompatible options, for example, when throwing dice, the chance that a six falls is the same as the chance that it will fall on the other five remaining numbers. Every single throw represents an incompossible set of potential throws that were not made. The Peircean notion of *chance* does not equal, however, a chaotic principle of a world in which no laws can be established. On the contrary, *chance* provides, for Peirce, the starting point for regularity, or *habit*, another important notion for Peirce's semiotics. Peirce uses both terms *chance* and *habit* as the driving forces of evolution. Peirce in his essay *The Logic of the Universe* even mentions the initial potentiality, the plurality of options that leads to the creation of habits.

> The evolution of forms begins or, at any rate, has for an early stage of it, a vague potentiality; and that either is or is followed by a continuum of forms having a multitude of dimensions too great for the individual dimensions to be distinct. It must be by a contraction of the vagueness of that potentiality of everything in general, but of nothing

> in particular, that the world of forms comes about. (Peirce CP, § 6.196)
>
> We see the original generality like the ovum of the universe segmentated by this mark. However, the mark is a mere accident, and as such may be erased. It will not interfere with another mark drawn in quite another way. There need be no consistency between the two. But no further progress beyond this can be made, until a mark will stay for a little while; that is, until some beginning of a habit has been established by virtue of which the accident acquires some incipient staying quality, some tendency toward consistency. (Peirce CP, § 6.204)

Peirce defines the evolution of the world in terms of diversity and uniformity, just as Kauffman (2000) talks about the subcritical state of the cells, which means the interface between constantly expanding diversity and freezinng in the established habits. Kauffman's theory was compared with H. Bergson's book *Creative Evolution* (Markoš 2003). I am of the opinion, however, that differently from Bergson, Kauffman shows points of contact with Peirce's philosophy of meaning formation by chance and habit, and with the definition of meaning as action in the world. Kauffman's description of the biosphere through the dissemination of diversity, incompatible with the second law of thermodynamics, which he presented in his book *Investigations*, is almost identical to Peirce's ideas expressed in the *Doctrine of Necessity Examined.* Kauffman defines the biosphere as continuously expanding its future possibilities, while retaining a reasonable degree of "traditionalism" (habit) when organisms act in well-established trajectories. Too many and too rapid changes could become malignant. Similarly, the natural language works (Kauffman claims that all of human society works in the same way). Language is constantly evolving[27], it is influenced by words from foreign languages, creates new rules, but still

---

[27] It has to be noted here that by "language" I do not mean the faculty of language as used by Chomsky. Obviously, the language faculty, according to Chomsky, is not evolving.

continues to follow the old ones. Changes allow semiotic systems to grow constantly, but too many changes are inconvenient. In practical and economic terms, it is more advantageous to keep up with the best footmarks, thus creating habits and a norm that help languages and species to exist. Living organisms transmit by heritability the accumulated rules of survival and orientation in the world to next generations, providing them with the right way to interpret the world. For Peirce-the pragmatist, such an important concept was consequently *habit*.

Habit is a notion regarding semiosis in general, not only the semiosis taking place on the lower threshold. An example of habitual interpretation in human cognition was proposed by U. Eco, inspired by a story about what happened to Marco Polo when saw a rhinoceros on his travels. In the book *Kant and the Platypus* (Eco 1997), Eco mentions how Marco Polo described in his travel journals an encounter with a unicorn, although somewhat blander and harsher than was described in contemporary books. Marco Polo, ignoring the existence of a creature such as a rhinoceros, in interpreting what he saw, used a category taken from his existing catalogue of mental representations (a unicorn). As C. Paolucci states, "for the sake of economics of labor, we tend to understand new meanings based on old unambiguous meanings" (Paolucci 2010, p. 382). Paolucci calls this process "interpretive comfort".

The notion of *habit* can be seen as a point of contact between Peirce's semiotics and structuralism. Hjelmslev established three notions replacing the Saussurean term *langue* in order to enlarge and redistribute the Saussurean terms of *langue* and *parole* that he found too limiting and unsatisfactory for a linguistic description, as they omitted the diachronic and communicative character of language. In his celebrated article *Langue et parole* (Hjelmslev 1942) he proposed the following three planes for the theoretical meaning of the term language:

> Scheme (language as pure form),
>
> Norm (language as material form) and
>
> Usage (language as a set of habits prevailing in a society)

Hjelmslev draws attention in this article to the misguided claim that language is the pure form (langue) that determines the linguistic act (parole). He claims, in contrast, that it is the linguistics act (usage) and the norm, which determine the scheme (the pure form in the sense of *langue*). A very similar observation was carried out by the Czech linguist J. Vachek in the essay *Can the Phoneme be Defined in Terms of Time?* (Vachek 1976). In this paper, Vachek stated that the linguistic form (langue) is derived from language use (parole) and not vice versa, thus disclaiming the apparent dogma of linguistic structuralism of the primacy of the form (langue) upon the meaning and upon the language use. This revised version of linguistic structuralist theory confirms the perception of the code establishment as a dynamic process, as is the case with Peirce, and as also described by K. Kull and others in the semiotic behavior of living organisms. The establishment of a semiotic system goes in a direction from the usage to the code, which implies that no predetermined forms exist (whether of language or living systems). Forms as such do exist, but are determined by usage, and are not eternally frozen and are connected to (and sometimes difficult to distinguish from) the notion of habit. The discussion of the primacy of the form concerns the process of protein folding as well. The forms of proteins (structures) are being studied (in the field of proteomics for instance), but the question of where the forms come from, has not been answered thus far. The application of Hjelmslev's belief, that the form is the result of usage (of a certain habit), to protein folding is plausible. From an evolutionary viewpoint, protein structures resulted as a response to environmental stimuli and were conserved in the following generations, becoming habits. The possibility that protein structures somehow preexisted in nature in the shape of Platonic forms (Denton, Marshall, and Legge 2003) is improbable, being rather results of various evolutionary factors such as fitness and thermodynamics.

> An understanding of the interplay of protein structure with both sequence evolution and functional/phenotypic evolution is necessary [...] To rigorously evaluate the possibility of a fold transition one would have to determine the viability of a series of mutations that connect the two folds. Both thermodynamics and kinetics of folding

> must be taken into account, as well as fitness effects due to function, all within in a context of population genetics. (Siltberg-Liberles, Grahnen, and Liberles 2011, pp. 749–752)

It is worthwhile remaining at this point with a fragment by Peirce:

> The evolutionary process is, therefore, not a mere evolution of the existing universe, but rather a process by which the very Platonic forms themselves have become or are becoming developed. (Peirce CP, § 6.194)

The present perfect and present continuous verb tense is an interesting choice by Peirce, presumably intentional, which effects the understanding of *forms* of life as constantly being under development, thus not predetermined and not working as blueprints of organisms, in our case one might say, blueprints of proteins. Despite the fact that he uses the term Platonic, his argumentation is distant from the classical understanding of Platonic forms.

There is a gradual passing from habits (usage) to frozen code (scheme), from the physical world to abstract forms, but of course, also in natural language. Similar to codes of organisms, even before the habits, before the creation of the form, there was a state of incompatible and undecided plurality of options. Since language is arbitrary, before the establishment of final rules, almost innumerable options for deciding the rules are possible. Just as with the example of the throw of dice, the final result is given by chance, one might say the final result is arbitrary. *Chance* in the case of code creating is different, however, from *chance* in the throw of the dice. While the former is arbitrary in the sense of being between numerous options, this particular was chosen, the latter is arbitrary in the sense that it is a question of pure coincidence. In other words, it would obviously be wrong to compare linguistic arbitrariness to a throw of a dice. The element that a dice-throw and linguistic arbitrariness have in common is an incompatible set of initial options. What the linguistic arbitrariness has in addition, however, is a causal relation between the initial plurality of options and the final choice of the option. This causal relation is guaranteed by linguistic usage, by habit, or by the principle of linguistic economy. Analogically, in biological codes, whether it be the genetic code, the protein folding code or the code of organism behaviour, amongst the initial plurality of options, there is a causality in relation to the final choice (the choice of the Final Interpretant). This causal relation is guaranteed, for instance, by evolutionary factors and by thermodynamic laws in the case of protein folding.

## 7.4 Arbitrariness and evolution

As I attempted to demonstrate in the previous paragraphs, the notion of arbitrariness in semiotics, which means the arbitrary nature of a sign (whether it be linguistic sign or a biological sign, for instance), does not prevent a causality between the two parts of the sign (signified and signifier or content and meaning). In natural language, there might be motivation for meanings of a certain word. It has been observed since ancient times that the connection between the meaning of words and their form is not necessarily completely arbitrary, Plato´s Cratylus for instance, see in (Cooper and Hutchinson 1997) and recent studies are demonstrating and even quantifying that in most parts of known languages the meaning of words from basic vocabulary is connected to a specific phonetic feature of the phonemes constructing the word, in other words, experimental studies demonstrating sound symbolism have been conducted (Blasi et al. 2016; Diatka and Milička 2017). This can be referred to as the casual relation between the content and meaning, although a causal relation does not prevent at the same time the arbitrary nature, which is guaranteed by the fact that there is no physical or chemical law governing the relationship between the two. The connection, even if motivated, is optional. As a consequence of this optionality, many languages in the world use different words to express the same meanings.

The same situation comes about when considering protein folding, or the genetic code. Relations in biological codes are not random, of course, or to put it in a different manner, are not deprived of casual explications. Causality in the genetic code, for instance, is due to evolutionary convenience or consummation of the energetic minimum. Nevertheless, the relationship between two parts consisting of the sign (amino acids and DNA bases) remains arbitrary, because the relationship between the two is not guaranteed by any physical or chemical necessity, it having been, originally, optional. Indeed, if the relationship was originally optional, it does not mean that it also continues to be optional after the establishment of the code. Once the code is established, its rules should be preserved. If the rules of natural language are not preserved, one cannot use language for communication. If the rules of the genetic code are not preserved, anomalies and pathologic cases are the result. This is, once more, the case of the Final Interpretant. Arbitrariness and conventionality are notions inseparable one from the other. Both arbitrariness and conventionality guarantee the existence of a code. Jacques Monod (Monod 1972) named the two inevitable parts of the genetic code as chance (*le hasard*) and necessity (*la nécessité*), chance referring to the optional character of the relation between the two parts of the sign, necessity referring to the conventionality of the code. Once the rules have been established, they should be conserved, because in the inverse case, the code could not be used to communicate and to pass on information to the following generations. Monod (Monod 1972, p. 90) even used the term *arbitrary* when referring to the function of allosteric enzymes, defining the function of enzymes as being *chemically arbitrary* ("chimiquement arbitraire"), but also inventing a new term to design this very nature of arbitrariness. He named the relationship between the function of a protein (enzyme) and its chemical design as a relation of *gratuity* (gratuité).

A. Markoš, in the Czech translation of Monod's essay, noted that the gratuity is a somewhat peculiar term, since in French, it means both "for free" and "unjustified", but in English and in Czech the term tends to be understood as "for free". Markoš claims, however, that Monod's term should be interpreted preferably as "unjustified", due to the fact that from an evolutionary standpoint, it is not right to talk about gratuity, but on the contrary, the price was quite in terms of how much energy was spent by previous generations. Thus, the synchronic "for free" state of chemical gratuity is counterweighted by the diachronic energy consumed during the evolutionary process[28].

> The trait seems to be for free from a synchronic viewpoint, i.e. from a viewpoint of living organisms. These no longer need to take into account the long ages of evolution, during which the enormous amount of energy – both physical and creative – was invested in the improvement of each trait. (Markoš 2008, pp. 177–178)[29]

I introduced briefly the notion of chance in Peirce and in Monod. I consequently presented the term gratuity in Markoš. I would now like do dedicate a few words to Darwin's use of the notion chance, specifically the notion of *random variation*, as it was at the very heart of his evolutionary theory. A. Rosenberg (Rosenberg and McShea 2008) already pointed to the misunderstanding of Darwinian theory by Modern Synthesis. Darwin used the term random variation, as he argued, because of the ignorance of the causation, not because of the non-existence of causation, as it was often misinterpreted.

> I have hitherto sometimes spoken as if the variations [...] had been due to chance. This, of course, is a wholly incorrect expression, but it serves to acknowledge plainly our ignorance of the cause of each particular variation. (Darwin 1987, p. 102)

Moreover, as Rosenberg noticed, the randomness of variation does not reside in mere ignorance of the prior causes, but resides, additionally, in the independence of a variation from the factors that determine its final adaptation (preserving the varied trait).

> The theory requires that in every generation heritable traits vary to some degree, and that this variation is

---

[28] Markoš focuses more on the term "gratuity" in the book *Evoluční tápání* (Markoš 2016).

[29] In the original "Zadarmo se vlastnost zdá být z hlediska synchronního, tj. z hlediska právě žijících organismů. Ty už nepotřebují brát v úvahu dlouhé věky evoluce, během nichž se do vylepšování každé vlastnosti investovalo obrovské množství energie – fyzikální i té tvůrčí."

> "random." […] The theory of natural selection does however rule out one cause of variation in heritable traits, namely a future cause in which new variation is guided by the needs of the individual who bears it. Indeed that is the major thrust of the word "random" in the phrase "random variation" in Darwin's theory. It is not that the appearance of a new trait is undetermined, that it is not fixed by prior causes. It is rather that the causes that fix it are independent of, unconnected with, the factors that determine its adaptedness. We say that variation is random "with respect to" adaptation. (Rosenberg and McShea 2008, p. 18)

Darwin even expressed a necessity, a *must-be*, of the causes in species variations.

> Whatever the cause may be of each slight difference in the offspring from their parents - and a cause for each must exist - it is the steady accumulation, through natural selection, of such differences, when beneficial to the individual, that gives rise to all the more important modifications of structure, by which the innumerable beings on the face of this earth are enabled to struggle with each other, and the best adapted to survive. (Darwin 1987, p. 131)

The variation that has occurred was caused by a certain cause in Darwin's theory, it not being random in connection with prior causes. Knowing the cause, however, of the modification is irrelevant to its effect, such as a better adaptation to the environment (or worse). We can therefore speak of randomness, of the chance of the variation with connection to its future consequences (not in connection with its prior causes). Hypothetically, other variations could have occurred to solve the situation with the same success. This is what is meant by the term *chance* in Darwin. The randomness does not mean, however, that there is not a causal relationship, it only means that the consequence of the variation (adaptation) is not necessarily linked to the very cause of the variation.

The aforementioned understanding of the term of random variation is related to the functional plasticity of organs. The polyfunctionality of organs, body shapes was discussed by Darwin in a detailed manner. He also detected a relationship between the particularity of shapes and their specific functions and the variability: the more an organ is particular, the more concrete function it has, the less it tends to vary:

> As long as the same part has to perform diversified work, we can perhaps see why it should remain variable, that is, why natural should have preserved or rejected each little deviation of form less carefully than the part has to serve for one special purpose alone, in the same way that a knife which has to cut all sorts of things may be of almost any shape, whilst a tool for some particular object had better be of some particular shape. (Darwin 1987, pp. 115—116)

As a protein example of what Darwin calls variability we could think of the already mentioned moonlight proteins which can vary in structure, that is, do not serve one special purpose, but are instead a kind of "universal knife". Other proteins, in contrast, are well specified in structure and consequently in function (or vice versa) and thus do not vary any longer. Haemoglobin's specific shape serves the specific function of transporting the oxygen molecules, the structure of haemoglobin is fixed, being formed as an assembly of smaller subunits of globular proteins arranged in a tetrahedral manner (a quaternary structure).

In summary, I tried in this chapter to approach the problematic biosemiotic question concerning meanings in the biosphere. I suggested comprehending meanings in terms of interpretation capacities, with the aid of Eco's notion of Space C and the notion of the phenomenal present. I consequently advanced by comprehending meanings in terms of code usage, that is, in terms of semiotic scaffolding as gradual creation of habits, codified rules which lead to Final Interpretants. I introduced the notion of scheme, norm and usage in Hjelmslev in order to distinguish between singular states of code creation. I commented on the relationship between arbitrariness and codification of codes both in natural language and in biology. Using the examples of the definitions of terms such as chance and necessity (Monod),

| | Arbitrary | Codified |
|---|---|---|
| Monod | Chance (gratuity) | Necessity |
| Darwin | Random variation | Adaption/Preservation |
| Peirce | Chance | Habit (Final interpretant) |

Table 7.2: Concepts of "chance" and "habit" as approached by Monod, Darwin and Peirce.

random variation and preservation (Darwin) or chance and habit (Peirce), I argued for a balance between what is arbitrary and what is habitual, both characteristics being fundamental preconditions for the existence of codes.

For Peirce, *chance* is not seen as a principle of a chaotic world in which it is impossible to lay down any laws. On the contrary, *chance* provides a random starting point in the formation of regularity and *habit*. Peirce defines the development of the states of the world in terms of diversity and uniformity, where there is a specific relationship of interdependency between these two terms. This leads us back to the already remarked initial potentiality, a plurality of options that guarantees the interpretation of Space C. Once the decision of choice is made, plurality disappears and habits can emerge.

> There is, therefore, every reason in logic why this here universe should be replete with accidental characters, for each of which, in its particularity, there is no other reason than that it is one of the ways in which the original vague potentiality has happened to get differentiated. (Peirce CP, § 6.209)

The coexistence of incompossible options gives rise to becoming a habit, which in turn transforms into a norm that can become a code (frozen rules). Once rules are established, there is a possibility to reinterpret them, although the Final Interpretant was decided. The change of the Final Interpretant is not particularly advantageous, one has to have good reasons to change the code rules, and more importantly, one has to receive consent from the speaking community. One can opt, for example, for a rule violation. I decided to use the word "apple" whenever I want to talk about weather. I will say, for example, "Today the apple is cloudy." The change in language rules will be disadvantageous for me from both social and communicative viewpoints. Theoretically, however, a situation could arise when it would be advantageous for me to change the code. The speaking community would accept my rules, they would start using them as well, it would become a *usage* (habit), thus a new *scheme* would be created. There was a celebrated case in Italy of the potential recognition of the new word "petaloso" two years ago. The word was invented by a pupil at primary school, which seemed appropriate to the concept he wanted to express, the properties of a flower that has many leafs, leafy. The teacher proposed the official recognition of the newly formed word to the Accademia della Crusca linguistic institute. Accademia della Crusca's answer was that the word was potentially recognizable, if its use could be proven. Thanks to social networks and media attention, the word "petaloso" has actually become very common.

To recapitulate, meanings, dynamic meanings in language and in the biosphere are being constantly shaped. The more robust, the more codified the code, the longer the tradition, the harder it is to change, yet the option is still there, the Final Interpretant may still enhance a new semiosis.

## 7.5 Semantics

In the previous chapter I focused on the notions of habit as a prerequisite of form (established code) and stated, following Hjelmslev, that the linguistic form, *scheme*, is determined by *usage* and not vice versa. In the context of organic codes, the situation is very similar. Behaviour habits or codes, instincts are automatic processes without an interpretative Space C, but this is only the synchronic state, and as was pointed by Markoš, the synchronic gratuity is always compensated for by a diachronic evolutive negotiation. Consequently, the present form (the behavioral *scheme* for instance) was preceded by the creation of habits and their usage. In the context of protein folding, the function of a protein, for instance the function (and consequently the structure) of an allosteric enzyme is, according to Monod and Markoš, a codified scheme defined by gratuity. Evolutionary negotiating, trying out possible structures (habits), etc. preceded the creation of this codified scheme of a given protein structure.

Once the habits are stabilised and the code is thus created (the scheme), it becomes useful to conserve the rules of the code so as to preserve it for future generations. For this purpose, not only the codification of rules, but also the writing system and the storage media were invented such as written symbols and paper for instance. By analogy, in the context of proteins, the genetic code and the order of amino acids in a peptide sequence guarantees the "written symbols" and the molecule of DNA guarantees the storage of the information. It is important to realize, however, that the written symbols are "only" a tool for storing code rules and for storing important information, important content. They are not themselves, however, the actual causation of the emergence of content.

The question I advanced at the beginning of this thesis was how to construe the grammar of a language, in particular, how to construe "a grammar for protein folding". I led my argumentation towards a statement that what is the most important to deal with when constructing a grammar is the asymmetrical relation between linearity and non-linearity, what can be, to some extent, conducted to the asymmetrical relation between the expression and content. I maintained that a grammar should be construed from the content and not from the expression, because of the participative relation that determines both and in which the intensive term is represented by expression and the extensive term is represented by content. As a matter of fact, grammars of natural language, construed from content, do exist and there have been several attempts to propose such a type of grammar. In the following paragraphs, I would like to comment on some of these grammars and would consequently like to discover what an application of this kind of grammar to protein folding might look like. In the grammars of natural language, this is the case of generative semantics theories for instance, or other semantic-based grammars. Before the application to protein folding, however, the question must be answered, as to what is meant by semantics here, since so many definitions of semantics occur in linguistics, depending on what kind of meaning is being considered. Certain linguists also add, for example, to the propositional meaning the so-called rhetorical meaning (Schmerling 2018). Meanings might be explicit but also implicit and in this manner semantics border on pragmatics.

The term semantics is used primarily to describe a linguistic discipline, which deals with the meaning of language expressions at its various levels, especially at the level of the meaning of singular words (lexical semantics). There are also broader meanings of the term semantics: semantics of programming languages, semantics in logic, Kauffman even speaks of semantics in physics (Kauffman 2000, pp. 109–111). Semantics can mean a wide range of phenomena that have very little in common. Kauffman speaks of semantics in physics in relation to the decision-making of autonomous agents. Semantics of programming languages, in contrast, do not allow any decisions or interpretations.

Within the semiotic theory of Ch. Morris, semantics is explained as opposed to syntax and pragmatics. A similar approach is presented by the Czech linguist Jan Kořenský, which divides the components of language into semantics, pragmatics and text-forming means. What is interesting, however, is the following formulation by Kořenský:

> The semantic basis is not, however, the only component of the non-expressive nature of the natural language. It is necessary to [...] formulate the other functional components of the given model. It is primarily the so- called pragmatic component of natural language.[30] (Kořenský 1984, p. 16)

This formula, apart from the distinction between semantics, pragmatics and expression, implicitly distinguishes between the expressive (expression related) and the non-expressive component of language (here semantics and pragmatics are classified), which is, in principle, a dis-

---

[30] In the original "Sémantická báze však není jedinou složkou přirozeného jazyka nevýrazové povahy. Je třeba formulovat [...] další funkční složky modelu. Je to především tzv. Pragmatikcá složka přirozeného jazyka".

tinction between expression and content. Kořenský presents a grammatical model from content to expression in which semantics and pragmatics stand together in opposition to expression.

Although the meaning of linguistic expressions has been dealt with since ancient times, semantics as a separate discipline began to form in the late nineteenth century. What was the meaning of "semantics" at that time, at the time of its origin? M. Bréal and his other contemporaries, V. Henry (Henry 1896) among others, began to study the meaning of words as a reaction to the later focus of linguistics on language development and phonetic laws. They lived in the period of Historical and Comparative Grammar. Linguistics of that time only dealt with expressions, ignoring the fact that words *meant something*. Bréal's argumentation (Bréal 1897) concerning the need to explore semantics is deeper than merely pointing out that linguists should pay attention to the meaning of words. His argument puts the word to the fore as a combination of expression and meaning. Bréal suggested that words cannot be considered regardless of their meanings, because the word becomes a word only because it *means something*, otherwise it would only be a sequence of sounds. Bréal maintained that the word is an inseparable connection between an expression and its content. It is consequently considered the predecessor of Saussure's notion of the linguistic sign as the connection between the signified and signifier. The very term *value* by Bréal is important in this regard. Bréal talked about the *value* of the sign as its meaning in connection with expression, this connection being given narrowly.

The original direction of the first semantic works was taken towards the need to describe language from a semiotic point of view, that is, to describe words as inseparable linkings of expression with content. This semantic project has failed thus far, however, since most semantic analyses of natural language, such as the one by Katz and Fodor (J. J. Katz and Fodor 1963) or the one by Langacker (Langacker 1990) deal with lexical meaning. These works therefore only study the meanings that are directly related to the lexical units. Said differently, such approaches take an expression unit (a word) and study its meaning(s), which means that they study meanings, but the direction of the study goes from expression content. This approach might lead to certain inadequacies, since some meanings have a more complex expression, which we do not identify when dealing exclusively with lexicon. The semantics of lexical meaning is nothing more than a processing from expression to content, and as I have tried to argue previously, this passage goes from an intensive term (expression) to an extensive term (content). As a consequence, some meanings might not be described, since the intensive term is a reduction of the extensive term. It is therefore not guaranteed that such a grammatical approach will describe all relevant content (meanings). Lexical semantics does not guarantee a semiotic conjunction of expression with content. It is an approach that goes from higher units (words or lexemes) to smaller units (morphemes and phonemes) and thereby diverges from classical grammatical description starting from smaller units and proceeding towards higher units. One cannot say from a semiotic standpoint, however, that it is an approach going from content to expression.

In order to keep Bréal's original definition of semantics and the required association of expression and content (meaning), there is a need for the grammatical description to move from the content to the expression (and not vice versa). Paolucci (Paolucci 2010) called this process *onomasiological*, and referred to this kind of semantics as *semiotic semantics.* As Paolucci claims, the content always precedes the expression in the semiotic function because "the language is used to express the contents, and not to articulate expressions, exactly as we sit because we are tired and not because our knees articulate". (Paolucci 2010, p. 372)

Paolucci's understanding of semantics emphasizes the impossibility of hierarchizing meanings, because the meanings are not unchangeable pre-existing language units, but are dynamic entities formed at each cut of the semiotic encyclopedia (fold). The fact that meanings are dynamic and therefore never exhaustively enumerable, implies that it is only possible to create a concrete grammar in every specific semiotic situation, at each cut of the semiotic encyclopedia. In every encyclopedic cut, a synchronic actual and local state of art might be described. Yet such a construing of specific grammar for every specific situation seems a very demanding task and unpractical and would be, in all probability, of no use. We are consequently looking for a more practical grammar from content to expression that would be able to describe dynamic meanings without enumerating them.

In order to create such a grammar from content to expression, Kořenský proposed a grammar "from a semantic basis" (Kořenský 1984). In the grammar proposed by Kořenský we must proceed from semantics, which is not limited to lexical meaning, but on the contrary, is based on meanings, regardless of lexical realization. Kořenský's method is based on an assumption that linguistic meanings can be expressed by a different expressive realisation, concretely by at least three basic strategies:

- lexically (by a single word)
- syntagmatically (by a syntagma)
- predicatively (by a proposition)

Consequently, there is no one-to-one correspondence between an expression unit and a content unit, the relation of correspondence is dynamic and is changing with every new cut of the semiotic encyclopedia. In other words, the relationship between expression and content is changing depending on the actual usage, depending on the actual speech.

Describing from content to expression is not so easy, which is also the reason why the majority of academic grammars are expression-based and not content-based. The most striking problems of content-based grammars have been presented by Faltýnek (Faltýnek 2013), who stated the counterarguments against the description from content to expression. He proposed a parody of content-based grammar to demonstrate the problems that a content-based grammer might lead to, based on the semantic analysis of Czech words:

(18) a. *lak* (a varnish),
b. *vak* (a bag) and
c. *tak* (thus),

Meanings of the three aforementioned words were defined by Faltýnek as follows:

(19) a. the material that covers the surface of objects, these objects are completely closed in it (a varnish),
b. a tool for storing the appropriate objects in a compact structure in its interior, which we can carry along with (a bag)
c. an anaphoric substitution as a summary of semantics of the previous assertions (thus)

After a comparison of the three meanings, Faltýnek concluded that all the three words share a common basal meaning that is "delimiting a certain content in a compact structure". Faltýnek consequently made use of the commutation test by Hjelmslev. The result of the commutation test showed that the three words express the same basic content, concretely "delimiting a certain content in a compact structure".

Therefore, according to the parody of content-based grammar by Faltýnek, the letters *l*, v, *t* are only variants of a single phoneme that can be interchanged without changing the lexical meaning. Obviously, this conclusion is false, sounds *l*, *v* and *t* are phonemes of the Czech phonological system (and not variants of one phoneme).

Faltýnek's example should serve as an extreme and parodic case of the impossibility and ridiculousness of a grammatical description based on content. Faltýnek's example does not show, however, a grammatical description from content to expression, but rather from expression to content. It indicates the direction of grammatical description from bigger units (words) to smaller units (phonemes), but both of the units are part of the expression plane, the content plane is considered only secondarily (after the choice of the units of comparison). This is an example of grammar based on lexical semantics. As I mentioned previously, grammars based on lexical meaning describe grammatical phenomena from expression to content, even though it might seem at first glance as a description from content to expression, since the standpoint of this description is the lexical *meaning*. This approach is actually based on a description of the meaning of lexemes and not of meaning as such. In other words, lexemes are expression units expressing meanings, but are not meanings by themselves. The arguments against a grammatical description from content to expression, as presented by Faltýnek, are not valid because they are not really referring to grammatical description from content to expression. It is only an illusion of lexical semantics.

If we want to see what such a description from content to expression might look like, always using Faltýnek's example, the actual description from content to expression would have to start from the content, thus would have to be based on the content of "delimiting some content in a compact structure", rather than starting from the singular lexemes expressing this content. This method is presented by Kořenský in his grammar model from content to expression. In Kořenský's example of the description from content to expression (Kořenský 1984, p. 21) we begin with the content of "a ruined medieval fortress" qualified as the start point of the analysis and from that content we derive possible expressions, concretely lexical, syntagmatic and predicative realisations:

(20) a. "a ruin" (*zřícenina*),
b. "ruined castle" (*zřícený hrad*),
c. "The castle is ruined" (*Hrad je zřícený*).

Faltýnek's example, with the words *lak*, *vak*, *tak*, is not a suitable counterargument for a grammar from content to expression. The example is just a leap-up procedure that is commonly used in language, and is a process from expression to content, and which, as I have already said, uses lexical semantics as a standpoint. Lexical semantics seeks to find contact points between different meanings of polysemic words such as, for example, "bachelor" or "ring". The problem is that the points of contact can always be found with a good imagination, the only real point of contact often being the common expression (word), so the words in the same category are found to have meanings far removed from each other, such as "old bachelor" and "seal" (J. J. Katz and Fodor 1963) or "ring" and "sporting arena" (Langacker 1990).

What do a ring and a sporting arena have in common? A ring circumscribes a finger circumference, a sporting arena limits the space for boxing matches. The words "ring" and "ring" delimit some content in a compact structure, exactly like a "varnish", a "bag" and the conjunction "so" in Faltýnek's examples. Faltýnek began his parody of grammatical description from a quasi-equal expression of three words (differing in one letter), Langacker started his description from a fully-equal expression (the terms “ring and “ring” are identical in expression). In fact, Langacker’s conclusion is in the end as absurd as Faltýnek’s, which was absurd on purpose. A sporting arena and ring are not alike either in terms of shape, the ring in the box has a square shape, in fact, in Italian even “quadrato” is used (literally “square”). This problem was approached by Paolucci (Paolucci 2010).

Even if it seems unintuitive, the aforementioned examples are examples of a grammatical description from expression to content and not vice versa. It is from expression to content because the description starts with a comparison of similarity of expressions (words).

Let me now try to outline the difficulties for the grammatical descriptions from expression to content in a different area than semantics. Take, for example, the categorisation of grammatical categories. The method remains the same as in preceding examples, that is, the following steps are made:

1. choose a set of expression units,
2. define respectively the meaning of every single chosen expression unit,
3. find such a part of meaning that every one of the chosen units shares,
4. if this is possible, then
5. categorise expression units into one category

Consider, for example, the following Czech words and try to categorise them in ine if the following manners:

(21) moje, tvoje, jeho
my, your, his

The words *moje, tvoje, jeho*, specify nouns, e.g:

(22) “moje kolo”, “tvoje šaty”, etc.
“my bicycle”, “your dress”, etc.

Thus, the words *moje*, *tvoje*, *jeho* belong to the category of adjectives.

(23) moje, tvoje, jeho
my, your, his

The words *moje*, *tvoje*, *jeho* refer to other elements in the text, they do not have a direct reference function, e.g:

(24) a. Karol vstává každé ráno brzy, vyžaduje to **jeho** náročná práce
‘Karol wakes up every morning early, **his** demanding work requires it’
b. Peter a **jeho** sestry se stěhují na venkov
‘Peter and **his** sisters are moving to the countryside’

The words *moje, tvoje, jeho* belong to the category of pronouns.

The definition of a grammatical category based on searching for common elements of words is arbitrary. Common elements can be found from more viewpoints and it involves choosing which aspect one considers more relevant. In Italian or English grammars, words like *my*, *your*, *his*, belong to the category of adjectives and rank among other adjectives such as *pretty*, *small*, *tall*[31]. In grammars of other languages, e.g. in Slovak or Czech, these words belong to the categories of pronouns and rank amongst other pronouns such as *I*, *we*, *the*m, *that.* We might consider other options for searching for the common characteristics for the words my, yours, his, and construe other grammars from expression to content.

I have introduced grammar problems from expression to content. I will now focus on issues of grammatical descriptions in the opposite direction. The problem of creating grammar from content is based, as Kořenský (Kořenský 1984, p. 13) states, in particular on the assumption that the content set is inconceivable, that is, its units are not empirically imminent, unlike expression units to which we have direct access. Kořenský adds, however, that the problem of the immediacy of the content set is merely a presumption, a prejudice, given that the expression

---

[31] Although these expression are more often classified at present within the category of determiners.

set units are equally empirically inconceivable as content set units, at the moment when we comprehend them in connection with the meanings. All sounds are empirically immediate, but sounds that express or delimit meanings (phonemes), are harder to detect empirically: where does the phoneme E begin and end? For every language elsewhere, many languages do not even have a phoneme E (in the sense of a unique phoneme E) but instead have several other phonemes, such as for instance a closed E, an open E or a nasal E.

Empirical immediacy of expression units is questionable. Kořenský further argues that grammars which reject a grammatical description from the content and only permit a description of the terms that are empirically imminent, i.e. any expressions (not necessarily expressions related with meanings), are problematic because they treat language only as one set, expression set, without any relation to the content set. Such approaches, one can say, only describe an "s-code". Grammars that reject describing the content set are based on the assumption that there are also contents which are not expressed in any way and are therefore considered impossible to be described by a grammar. This assumption is unacceptable, however, and improper in terms of a semiotic standpoint. From a semiotic standpoint, there are only meanings associated with expressions, other meanings simply cannot be imagined only if these are meanings not belonging to the semiotic description. According to Kořenský, meanings in the language are standardized in the same way as expressions are standardised. In other words, meanings are codified, normalised by *usage* and *scheme*, thus, are grammatically describable and the normalization of meanings is a prerequisite for language use in communication:

> Without the existence of this dynamic, but normalized set of relational models of the "states of the world", that is to say, without a set of stable rules of depiction of reality in language, communication is just as impossible as it is

> without dynamic but normalized principles of use of the means of expression. (Kořenský 1984, p. 14)[32]

Linguistic contents do not represent all possible states of the world, we cannot, due to our sensory and cognitive limits, think of all objects in the world at all angles and with all the smallest particles that constitute them. We perceive the world as it is useful to us. As Faltýnek (Faltýnek 2013) points out, living organisms can only perceive the world from given angles, and for each species these states can be different. We can smell, see and hear only certain states of the world, and we segment them in a certain way, which is given by our physiological structure. We form semiotic sets, the systems we use to orientate and communicate in the world. Based on these sets, a conventional code is constructed, assigning meanings to expressions. This is happening locally, based on the given cut of the semiotic encyclopedia. Kořenský stated that there are consequently no meanings, not in language, that are not related to expressions. Since the meanings are constantly changing, however, given that also the extralinguistic reality is changing, it is impossible to enumerate them in general terms. One can only enumerate meanings locally, for the local cut of the semiotic encyclopedia.

Let me use one more example by U. Eco (Eco 1997). When Marco Polo sees what he sees, he interprets it as a unicorn. An aboriginal who stands nearby and looks in the same direction as Marco Polo does, sees a completely different animal, he sees what we call a rhinoceros. Marco Polo and the aboriginal are equipped with different encyclopedic knowledge. It is therefore impossible to set up a set of minimum content units. General meanings do not exist, meanings only exist locally.

J. Kořenský, when defining the term of the "semantic basis", argues with a similar assumption: the semantic basis is not an unchangeable

---

[32] In the original "Bez existence tohoto dynamického, ale normovaného souboru relačních modelů 'stavů světa', tj. bez ustálených pravidel jazykového zobrazování skutečnosti, je komunikace stejně nemožná jako bez dynamických, ale normovaných zásad užití výrazových prostředků jazyka."

a priori existing unit, it is a historically creating dynamic result of interactions (Kořenský 1984, p. 16).

## 7.6 A content-based protein grammar

In the previous section I explained what I mean by content-based grammar. I am taking into account the grammatical model by Jan Kořenský which comprehends meanings as dynamical entities, impossible to enumerate in general terms, yet always definable and enumerable by a local cut of the semiotic encyclopedia. The core idea of this grammatical model is to describe language from content to expression, that is, to describe such expressive units that are responsible for expressions of meaning. The advantage of this grammatical model is, regardless of the expression, to differentiate meanings that are different and not to differentiate meanings that are identical. This observation seems intuitive and not even worth mentioning, nevertheless the fact is that some grammatical descriptions fail in such a task. Consequently, using the model by Kořenský,

(25) a. *ring* as a piece of jewelry and
b. *ring* as a sporting arena

have semantically two different meanings. However,

(26) a. *ruined castle* and
b. *a ruin*

express semantically one and the same meaning.

Such an approach, when applied to the grammatical categories, obviously leads to a reclassification of basic grammatical categories, just as Kořenský did in the Czech grammar *Akademická mluvnice češtiny II* (Komárek and Kořenskỳ 1986). In *Akademická mluvnice češtiny II*, Kořeský proposed only four basic parts of speech rather than the classic number of ten (substantives, adjectives, verbs and adverbs). The reason for this reduction of the number of basic parts of speech was a functional content-based approach towards grammatical categories. This approach classifies grammatical categories functionally rather than morphologically, that is, categories are defined by sharing the same functional properties, by expressing a certain kind of content and not by sharing the same morphological properties (i.e. flexion, etc).

Could the approach by Kořenský be applied to protein folding? A description of proteins from content to expression could be understood as describing proteins from non-linearity to linearity, in other words, from structure to sequence. This kind of describing of proteins I already prefigured with my thought experiment in the first part of this thesis. The like methods in protein studies were also recently proposed in the field of protein studies, I will comment briefly on them in the following paragraphs. Passing from the protein structure to the peptide sequence might, however, be quite similar to the methods of approaches based on lexical semantics, such as the one by Langacker. The direction from higher units to smaller units does not automatically equal the direction from content to expression.

The studies proposing an alternative approach in proteomics, that is, to start the description from structure and not from sequence, are not all that common and are rather recent. Attention directed towards changes in the optic in proteomics is providing important observations, pointing to the non-unequivocal correspondence between sequence and structure. One of the important discoveries given by this approach is a possible rethinking of the hierarchisation of the protein folding process. Over recent decades, in fact, different hierarchical units are more preferably studied in comparison with secondary and tertiary protein structures. These units are represented by so-called protein domains, structural motifs, and even units such as supersecondary structures (Kister 2015), minimal functional cores or minimal structural cores (Holm and Sanders 1997) occur. The aforementioned protein units represent a kind of transitional state between secondary and tertiary structures. Structural motifs are spatial arrangements of structures that are responsible for the connectivity between the secondary structures of a protein, and are not characterized by a specific function, that is, the same structural motifs can be found in proteins with different functions. Protein domains are, in contrast, semiindependent structures, which might fold and evolve independently from the rest of the protein.

The term supersecondary structure (SSS) was proposed by Kister (Kister 2015) and is a perfect example of describing proteins from structure to sequence. Kister, in contrast to homology modelling methods, suggests starting from a set of known structures, defining the folding places and only afterwards observing the sequence, that is, seeing what part of the peptide sequence is related to the folding places. This method was designed as a reaction to the gaps in homology modelling, based on studying homology in "atomic structures", that is, homology in peptide sequences. According to Kister, homology modelling fails to some extent because of the fact that not all the sequence is responsible for the folds, thus, even a high percentage of sequence homology might lead to different structures. In order to avoid this inaccuracy, he suggests only defining those parts of the peptide sequence that are responsible for the folds. To arrive, however, at such results, only the direction from structure to sequence is possible.

> There are several reasons for using SSS rather than the atomic structures of proteins in our study. First, SSS of a protein defines arrangement of secondary structure elements in space and has a strict and unambiguous definition [...] Secondly, proteins with an identical SSS may belong to different families, have diverse protein functions and hence, possibly, very little global sequence similarity. Proteins with similar SSS, but widely dissimilar primary sequences, are the most informative for discovering residue distribution rules. A further advantage of SSS is that it simplifies protein modeling analysis. Use of "protein skeletons"—SSS-rather than full-bodied atomic three-dimensional protein structures—allows one to uncover the common structural features that are specific and unique to a given fold. (Kister 2015, p. 45)

Kister's method, although proposing a direction from structure to sequence, does not take into account the very functional similarities in proteins, that is, the protein function (content) remains irrelevant for this method. An approach also considering the functional aspect was proposed by Holm and Sanders back in 1997. The term "minimal functional core" was introduced. Minimal functional cores were defined as:

> Regions which are invariantly conserved in all members of a superfamily of functionally and structurally related proteins. In lysozymes, the minimal functional core consists of a small b sheet located on one side of the substrate and two a helices at the back of the substrate. It seems plausible that these elements are the minimal set required to sustain substrate binding and catalysis. The minimal functional core of lysozymes does not have the characteristics of a structural unit that would be capable of folding independently. (Holm and Sanders 1997, p. 170)

The minimal functional units, as exemplified by the example of the lysozyme enzymes family, are lower level structures than protein domains and might not be able to fold independently. This observation leads to the conclusion that functional subunits are not necessarily structurally independent units. In other words, a minimal functional core is probably a plausible adept for an analogy to Kořenský's semantic basis. A similar approach was proposed by Berezovsky and the "closed loops" or "functional loops" as minimal functional protein units (Berezovsky, Guarnera, and Zheng 2017).

(27) a. *ruin*
b. *ruined castle*

are two examples of the expression of one and the same semantic basis which are not capable of "folding" independently, that is, are not able to function independently: normally they are only one part of an utterance. Only the predicative realisation

(28) *The castle is ruined*

can be considered an autonomous "fold". The question remains as to whether the units of a minimal functional core are really analogic to a content-based approach. As a matter of fact, a *minimal func-*

*tional uni*t corresponds to a *minimal structural unit*, according to Holm and Sanders, which implies that the analysis began with an already-made correspondence between function and structure, but Kořenský's method requires the analysis to begin from content (function) independently from structure (expression). The method presented by Holm and Sanders, the method of searching for minimal functional units, deals with protein functions, one the one hand, but only deals with functions already assigned to concrete proteins, on the other. This method is therefore not a real content-based approach as presented by Kořenský, it is an expression-based method of classifying proteins by means of finding the smallest functional entities which are at the same time structurally identical. Yet, similarly as in case of the pronoun/adjective class in grammatical categories, another possible classification is plausible here. Indeed, Holm and Sanders' study was challenged by Kobe (Kobe and Kajava 2000) who argued with the existence of solenoid proteins, which represent a specific class of proteins that does not correspond to the classification based on protein domains or supersecondary structures. To put it in a different manner, there is such a class of proteins that share important structural (geometrical) similarities, but these similarities are of a different order than secondary structure classes.

> The hierarchical classification schemes, exemplified by the databases, distribute protein structures into trees with the highest level of classification described by a "class" […] The solenoid proteins, despite their similarities, are split into different classes by these hierarchical classifications […] Structural similarities are indeed detected by geometric comparisons among proteins belonging to many different solenoid classes. The solenoid fold therefore appears to represent a novel grouping of structurally related proteins, where the similarities are often detectable by geometric comparisons, but where classification according to secondary structure or even minimal cores becomes inappropriate. (Kobe and Kajava 2000, p. 514)

What is more interesting is the fact that Kobe argues that solenoid proteins not only share common structural properties (repeats arranged in a superhelix manner) but also share functional properties:

> Their nonglobular shape appears to be one of the principal reasons for most of the solenoid proteins to share a functional property; that is, involvement in protein–protein interactions (Kobe and Kajava 2000, p. 511).

Interestingly enough, this directs me to the deduction that both types of classifications, that is, both "supersecondary structures-based" and "solenoid structures-based" classifications of protein structures assign a specific function for specific structures. In a like manner, the example of the Czech words *lak*, *vak*, *tak* can be classified as belonging to both the structural and functional same category.

I briefly introduced some of the existing studies treating proteins from structure to sequence. I argued that these studies might, nonetheless, diversify in classification of different protein types. An analogy to Kořenský's method of the content-based grammatical description is difficult to apply to protein studies, given the fact that thus far, the field of proteomics is only in its beginnings. A number of approaches exist to dealing with proteins and are often incompatible between themselves. It could be fruitful for future research to consider a content-based protein grammar that would start the description from the protein function independently from the structure. This approach might lead to a reclassification of protein types, e.g. some kinds of moonlight proteins might not be considered as one and the same protein, rather, having two different functions they would be classified as two different proteins. Similar approaches have led to two main conclusions thus far:

1. Higher units, such as the supersecondary structures (Kister 2015), are more important than singular amino acids because are directly responsible for the folds, in consequence for protein function.

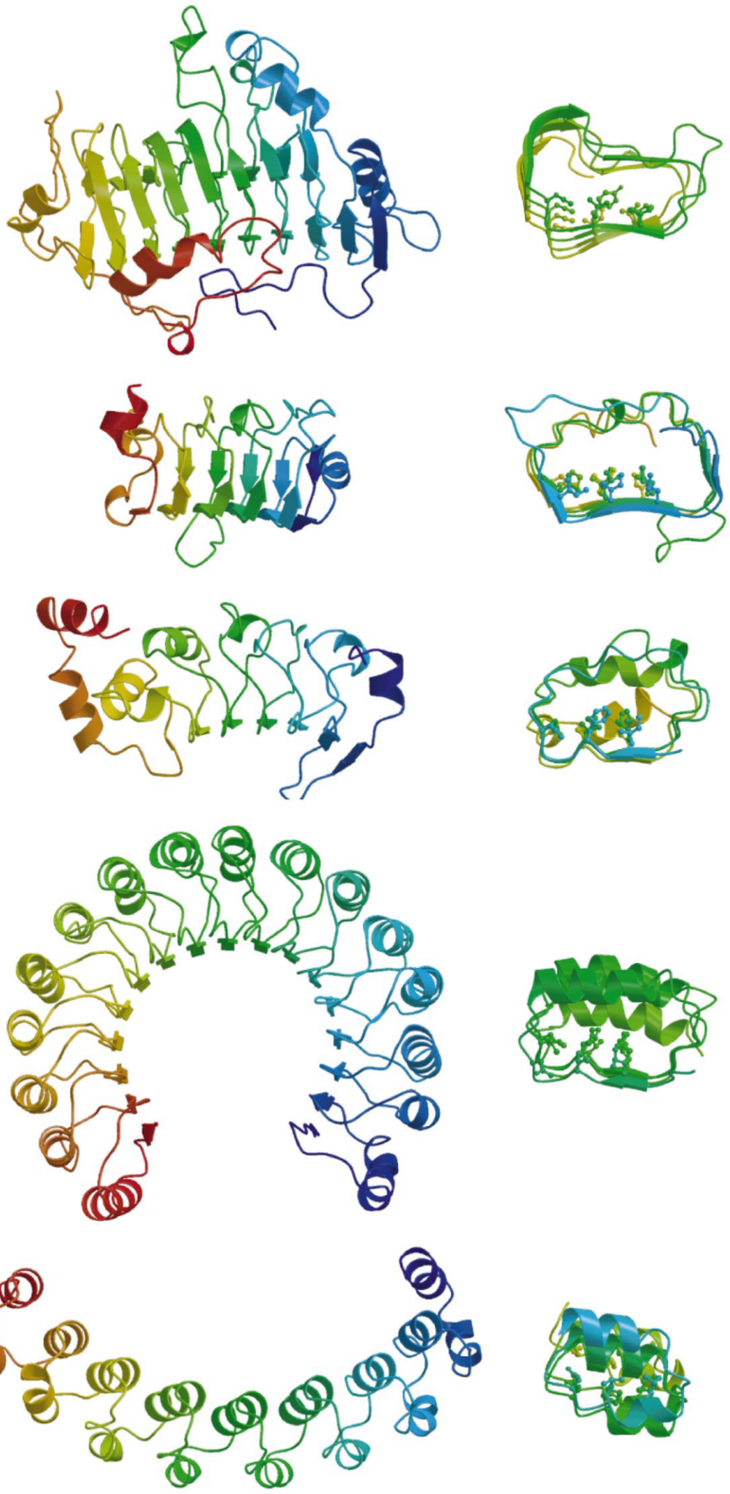

Figure 7.1: Selected solenoid structures (Kobe and Kajava 2000, p. 512).

2. The existing categorisation of protein subunits might not be the only possible and it can lead to imprecisions in categorising of proteins (Kobe and Kajava 2000).

As was demonstrated in this chapter, protein classification is just as arbitrary as classification of grammatical categories. Many options for classifying protein subunits are equally plausible. To avoid inconsistencies and to unify the viewpoint on what a minimal unit is, the method presented by Kořenský would be a possible solution in protein studies. In other words, proteins might be studied starting with the function description, not with the similarity in structure. After defining the function, various structures expressing this function would be detected. These structures would consequently be analysed in terms of the folding places definition (as proposed by Kister) and the sequences responsible for the folding points would be studied in the final phase of the analysis. Such an approach towards protein studies can be comprehended as an analogy to content-based grammar in natural language. As I argued previously, not all content-based grammars satisfy the requirement to deal with "semiotic semantics". One of the content-based grammars that satisfies the requirement of "semiotic semantics" is the grammar proposed by Kořenský. This kind of grammar deals directly with meanings (and not with lexical meanings). Analogically in content-based grammar of protein folding, protein functions should represent the starting point of the analysis. Thus far, this kind of approach is not a matter of practice in protein studies. Some approaches dealing with categarosation of protein subunits based on function (Holm and Sanders 1997) are close to my proposal, yet they only deals with already existing relation between function and structure. In other words, they do not start their analysis from the function, but from the function connected to a given structure, which is, so to say, a process of description analogic to Faltýnek´s parodical example.

# 8

# Conclusions

The inspiration for this thesis was found in some works on protein folding that analogise the process of the folding of a peptide chain to grammar, or even more concretely to syntax in natural language (Searls 2002; Gimona 2006; Kister 2015). The reasons for this analogy are at least two.

Fisrtly, the whole process of protein synthesis is based on codes, whether we speak about the DNA-to-peptide chain translation or the consequent peptide chain folding. These codes are a result of evolutionary negotiation and are characterised by arbitrariness and conventionality, similarly to natural language. To put it more clearly, chemical or physical explanations are not sufficient to unveil satisfactorily the process of protein synthesis, thus, an explanation of a different rank was needed and some scholars applied a linguistic point of view in order to understand better the process of protein synthesis.

Secondly, a grammar of a language (or syntax) is conceived by these studies as a description of obtaining a more complex structures from less complex structures (generating of words or sentences from phonemes or morphemes for instace). By analogy, one can assume a protein grammar such a description that explains the passage, a generating of a complex structures (proteins or protein subunits) from basic structures (nucleobases or amino acids).

I tried to conceive this analogy in a more complex way, that is, I proposed to consider a protein grammar in a way that is not related to the understanding of syntax as a lego-like building mechanism, rather is a way of explaining the complicated relation between linearity and non-linearity. Additionally, I proposed to consider a grammatical approach that is directed in an opposite way, that is, that starts with more complex structures and continues to the basic structures. Several notions related to the relation between linearity and non-linearity were commented and consequently applied to the protein folding process: teridentity relation by Peirce, participative opposition by Hjelmslev, double articulation by Martinet and other notions taken from linguistics and semiotics. I hope that the reader don´t get confused by the plurality of schools of thought reffered to in this thesis, such as European structuralism and Peircean semiotics. I believe that despite all incongruities between singular school of thought, all semiotic and linguistic description has to deal with the tension between linearity and non-linearity.

The scope of the analysis I proposed was to reconsider the existing approaches in proteomics, by linguistic analogy, in light of a content-based grammar. Most of the current methods in protein studies are based on finding the final protein conformation with aid of only informations given by the string of amino acids (homology modelling). With an argument that such approaches might be (and often are) limited due to the fact that when reducing non-linearity to linearity, some information get lost, I proposed to start the description from the other side, from already folded protein. This is not to state, however, that homology modelling is a wrong scientific method. My observation is simply that it is opportune to combine both approaches, the ones based on analysis of strings and the others based on analysis of the protein functions. Exactly like the dictionary model makes part of the encyclopedic model, a study of linear peptide strings has obviously an important place in the protein studies.

To conclude, the analogy between protein folding and grammars of natural language can be fruitful for both linguistics and biology. Back in 1995, Mantegna´s application of linguistic laws to DNA strings helped to the later discovery of the non-coding DNA functions. In this thesis, it was exemplified how in both cases of linguistics and biology, the grammatical description from content to expression reveals to be an exhaustive and efficient grammatical approach. I hope that the change in the viewpoint on protein folding proposed in my thought experiment may stimulate new ways to study proteins.

# Summary in Czech language

Proteiny jsou makromolekuly sestavené ze zřetězených aminokyselin kódovaných genetickým kódem. Skládání proteinů, proces získávání konečné proteinové struktury původního peptidového řetězce, je dosud velmi obskurní, pravidla určující skládání proteinů teprve čekají na objevení. Znalost chemických a fyzikálních vlastností peptidového řetězce, který je potenciálně skládan, tolik nepomůže při pochopení otázky *Proč se proteiny skládají tímto způsobem a nikoli jiným?* Chemie a fyzika hrají roli, přesto se zdá, že nejsou rozhodující odpovědí, hledá se vysvětlení jiné podstaty: “proteinový kód” nebo “proteinová gramatika” nebo “proteinová syntax” jsou pojmy, které se vyskytují. Tato situace je velmi podobná dešifrování genetického kódu, protože se předpokládá, že skládání proteinů je mechanismus, který byl získán přirozeným výběrem, podobně jako genetický kód. Tato disertační práce se pokusila propojit lingvistické teorie a studie o proteinech, což umožnilo další možný překlenutí lingvistiky a biologie.

# Summary in Italian language

Le proteine sono macromolecole costituite da amminoacidi concatenati, codificati dal codice genetico. Il ripiegamento delle proteine, un processo per ottenere la struttura proteica finale della catena di aminoacidi originale, è ad oggi un meccanismo molto oscuro. Le regole che governano il *folding* delle proteine non sono state completamente scoperte. La conoscenza dettagliata delle proprietà chimiche e fisiche degli aminoacidi non aiuta molto a capire la domanda *Perché le proteine si piegano in questo modo e non in un altro*? La chimica e la fisica hanno un ruolo importante, ma sembrano non fornire la risposta cruciale. Una spiegazione di un'altra essenza è stata richiesta: "un codice proteico" o "grammatica delle proteine" o "sintassi proteica" sono termini che occorrono negli studi biologici (Searls 2002; Gimona 2006; Kister 2015): un meccanismo piuttosto convenzionale, come un codice, viene richiesto piuttosto che spiegazioni puramente chimiche. Questa situazione è molto simile alla decifrazione del codice genetico, in quanto si ritiene che il ripiegamento delle proteine sia un meccanismo ottenuto dalla selezione naturale, in modo simile al codice genetico.

In questa tesi ho cercato di consolidare teorie linguistiche e studi sulle proteine, fornendo un possibile ponte tra linguistica e biologia. L'interfaccia tra queste materie è oramai abbastanza stabile, soprattutto nella terminologia biologica. Termini come *traduzione*, *traslazione*, *codice genetico* sono stati presi dall' ambito linguistico, e le analogie tra il linguaggio naturale e il codice genetico non sono state poche. In questa tesi, un'analogia viene fatta invece con il ripiegamento delle proteine, sviluppando così le idee già accentuate da alcuni studiosi (Gimona 2006; Kister 2015).

Si può dire che la biologia molecolare, e in particolare la proteomica, è in qualche modo correlata allo strutturalismo – nel senso dello strutturalismo linguistico del secolo scorso, che ha tenuto il favore della forma sulla sostanza –. Le proteine hanno la loro funzionalità metabolica grazie alle loro forme o strutture spaziali, la realizzazione materiale essendo di minore importanza. Allo stesso modo, le lingue sono – in termini di linguistica strutturale – strutture di relazioni tra elementi, completamente dissociate dalla loro realizzazione materiale. Per dare un esempio, consideriamo la traduzione interlinguistica, dove una parola può essere tradotta in varie lingue: così, lo stesso significato linguistico può essere mediato in diverse lingue grazie all'indipendenza della forma e della sostanza linguistica. Lo stesso principio è valido per una proteina e la sua funzione. La funzione delle proteine è riconosciuta esclusivamente dalla sua struttura. In virtù di questa proprietà, negli anni settanta del secolo scorso i ricercatori hanno iniziato a produrre chimicamente proteine artificiali che hanno un design chimico completamente diverso rispetto alle proteine naturali, ma che presentano le stesse relazioni strutturali. Le proteine sintetiche hanno il potenziale per essere utilizzati per scopi farmaceutici.

Il problema irrisolto nella ricerca attuale negli studi sulle proteine è il processo di transizione dalla sequenza lineare di amminoacidi (la catena proteica) a una conformazione tridimensionale unica. In effetti, una singola sequenza di amminoacidi può formare strutture differenti in diverse proteine e al contrario, una struttura proteica identica può essere derivante da sequenze diverse. Sorge inevitabilmente la seguente domanda: perché una sequenza si piega in un modo in una situazione e in un'altra in un'altra situazione? O in altre parole: "qual è la relazione tra sequenza e struttura?". Una delle risposte a questa domanda riguarda i fenomeni biologici legati alle cosiddette modificazione epigenetiche, laddove l'ipotesi epigenetica ritiene che non solo le informazioni contenute nel DNA, ma anche le influenze esterne – come ad esempio la temperatura ambientale e il contatto con altre sostanze – creino il prodotto finale della sintesi proteica.

Per capire meglio il rapporto tra la sequenza di aminoacidi e la struttura finale, ho proposto l'analogia con il linguaggio naturale. Anche negli studi linguistici, una delle domande principali riguarda la relazione tra una catena lineare, il *suono* e il contenuto linguistico non-lineare, il *significato*. Nei termini di Peirce, questa relazione riguarda il rapporto tra diadicità e triadicità.

Nell'esperimento mentale proposto nella tesi, una catena di aminoacidi potrebbe essere, per l'interpretazione semiotica dei fenomeni biologici, compresa come una linea continua, massa non articolata o possibilità indecisa. La catena degli aminoacidi è un "corpo senza organi". È un "corpo senza organi" a meno che non sia piegato. Una volta che si piega, il continuum viene spezzato dall'azione di piegamento e di diadicità che emerge; le pieghe sono punti di frontiera tra triadicità e diadicità. Di conseguenza, possiamo parlare delle relazioni diadiche come di espressione attraverso cui il contenuto è codificato. Possiamo parlare di amminoacidi come unità discrete di espressione organica che codificano per il contenuto organico rappresentato dalla funzione di una proteina, o dalla forma della proteina, poiché una struttura proteica e la sua funzione sono nozioni coestensive.

Nel mio esperimento mentale ho suggerito di pensare agli amminoacidi non come unità iniziali che generano le strutture proteiche, ma come una massa amorfa che solo nel momento di piegamento dà nascita alle singolarità. Poiché le informazioni contenute nelle sequenzialità di amminoacidi non dicono niente sul fatto come andrà a essere piegato, la "sintassi delle proteine" non è stata scoperta. Solo una volta piegata possiamo individuare i punti del piegamento della catena di aminoacidi. Come conseguenza, non è giusto dire che gli amminoacidi *generino* le proteine. Contenuto ed espressione sono due piani separati riuniti grazie all'abitudine e alla convenzione. Naturalmente, possiamo stimare quale sia il significato di una determinata espressione, ma non è la stessa cosa che significa il termine *generare*. Inoltre, possiamo stimare quale sia il significato di una determinata espressione solo se conosciamo le regole di espressione, cioè solo se conosciamo la grammatica. In altre parole, solo se sappiamo in che modo all'interno di una determinata unità di espressione del codice si riferisce a un'unità di contenuto, possiamo stimare quale sia il vero significato.

Possiamo tracciare un'analogia fonologica per una comprensione migliore. Quando si distingue tra i fonemi /b/ e /p/, tra la consonante sonora e sorda, l'unica cosa che conta è il punto di ripiegamento della continuità del suono. Quando ho bisogno di distinguere tra le parole *palla* e *balla*, l'unico punto distintivo è la linea di piegatura del suono, tutto il resto è il materiale fonetico puro che non ha un ruolo nel distinguere il significato. Ma come facciamo a sapere dove inizia il puro materiale fonetico e dove finisce? Possiamo solo saperlo una volta che conosciamo il significato delle parole italiane *palla* e *balla*. I fonemi hanno senso solo perchè differenziano i significati. È il contenuto triadico che si potenziava per creare la diadicità dell'espressione. Fino al collegamento con il livello di contenuto, il livello di espressione non era diadico, era semplicemente continuo. Per questo motivo ho proposto di cambiare l'ottica del cosiddetto dogma della Sintesi Moderna di Darwin che presuppone unidirezionalità dal DNA alle proteine. Il punto di vista che suggerisco va nella direzione opposta, quindi dalle proteine alle amminoacidi e infine al DNA. Ovviamente questa è una direzione della descrizione teorica e non dei processi biologici, che vanno, senza dubbio, dal DNA alle proteine (nonostante il fatto che con le scoperte epigenetiche anche l'unidirezionalità biologica viene rivalutata).

Per riassumere, abbiamo il continuum di una catena di amminoacidi, abbiamo la linearità di una catena di amminoacidi e abbiamo la connessione con il contenuto (funzione della proteina) come un punto di piegatura che rende la transizione dalla triadicità della proteina alla diadicità di una catena di amminoacidi articolata (piegata). La descrizione del processo biologico dalla proteina alla catena di amminoacidi potrebbe essere paragonata con dei metodi linguistici che cercano di descrivere il linguaggio naturale del contenuto all'espressione. Come modello di una descrizione linguistica tale ho preso la grammatica di Jan Kořenský (Kořenský 1984). Con questo modello ho cercato di trovare applicazioni possibili nel ripiegamento delle proteine. Ho esteso le analogie esistenti tra il ripiegamento proteico e il linguaggio a una possibile "grammatica del ripiegamento proteico".

Nonostante diversi studi interdisciplinari linguistico-biologici, una considerazione della grammatica dal contenuto all'espressione non è mai stata proposta in precedenza. Credo che un approccio di questo genere potrebbe essere valido negli studi sulle proteine, basandomi sul fatto che alcuni biologi hanno già proposto di guardare le proteine da un punto di vista più ampio: non focalizzando sulla catena di amminoacidi come punto di partenza,ma cercando di partire dalla proteina già piegata verso i punti delle piegature (Kister 2015).

Lo scopo dell'analisi che ho proposto era di riconsiderare gli approcci esistenti in proteomica, per analogia linguistica, alla luce di una grammatica basata sul contenuto. La maggior parte degli attuali metodi negli studi sulle proteine si basa sulla ricerca della conformazione finale della proteina con l'aiuto solo delle informazioni date dalla catena di amminoacidi (*homology modelling*). Con l'ipotesi che tali approcci potrebbero essere (e spesso sono) limitati a causa del fatto che quando si riduce la *non linearità* alla *linearità* alcune informazioni si perdono, ho proposto di iniziare la descrizione dall'altra parte – dalla proteina già piegata. Ciò non significa tuttavia che i metodi dell'*homology modelling* siano metodi scientifici sbagliati. La mia osservazione è semplicemente che è opportuno combinare entrambi gli approcci, quelli basati sull'analisi delle stringhe e gli altri basati sull'analisi delle funzioni della proteina. Esattamente come in Eco (Eco 2007) il modello di dizionario fa parte del modello enciclopedico, uno studio di catene di amminoacidi lineari ha ovviamente un posto importante negli studi sulle proteine.

Per concludere, l'analogia tra il ripiegamento delle proteine e le grammatiche del linguaggio naturale può essere fruttuosa tanto per la linguistica quanto per la biologia. Nel 1995, l'applicazione della legge di Zipf alle stringhe di DNA (Mantegna et al. 1995) ha contribuito alla scoperta successiva delle funzioni del DNA non codificante (junk DNA), ciò che prova che una ricerca interdisciplinare tra biologia e linguistica, oltre semplice analogie o metafore, può portare i risultati importanti. In questa tesi è stato esemplificato come in entrambi i casi di linguistica e biologia, la descrizione grammaticale dal contenuto all'espressione rivelasse un approccio grammaticale esaustivo ed efficiente. Spero che il cambiamento del punto di vista sul ripiegamento delle proteine proposto nel mio esperimento mentale possa stimolare nuovi modi riguardo al modo di studiare le proteine.

# Bibliography

Alberch, P.

1991 "From genes to phenotype: dynamical systems and evolvability", *Genetica*, 84, 1, pp. 5-11, DOI: 10.1007/bf00123979.

Ambrosio, C. and C. Campbell

2017 "The Chemistry of Relations: Peirce, Perspicuous Representations, and Experiments with Diagrams", in *Peirce on Perception and Reasoning: From Icons to Logic*, ed. by K. Hull and R. Atkins, Routledge Studies in American Philosophy, Taylor & Francis, pp. 86-107, ISBN: 9781315444628.

Andras, P. and C. András

2005 "The origins of life - The 'protein interaction world' hypothesis: Protein interactions were the first form of self-reproducing life and nucleic acids evolved later as memory molecules", *Medical hypotheses*, 64 (Feb. 2005), pp. 678-88.

Barbieri, M.

2003 *The Organic Codes: An Introduction to Semantic Biology*, Cambridge University Press, ISBN: 9780521531009.

2007 "Is The Cell A Semiotic System?", in *Introduction to Biosemiotics*, Springer Netherlands, pp. 179-208, DOI: 10.1007/1-4020-4814-9_8.

Barbieri, M.

2011 “A Mechanistic Model of Meaning”, *Biosemiotics*, 4, 1, pp. 1-4, ISSN: 1875-1350, DOI: 10.1007/s12304-010-9103-z.

2012 “Organic Semiosis and Peircean Semiosis”, 6 (Aug. 2012).

Bellini, D. and S. Schneider

2003–2018 “Banca dati dell’italiano parlato (BADIP)”, in Graz: Karl-Franzens-Universität Graz.

Berezovsky, I. N., E. Guarnera, and Z. Zheng

2017 “Basic units of protein structure, folding, and function”, *Progress in Biophysics and Molecular Biology*, 128, pp. 85-99, ISSN: 0079-6107, DOI: 10.1016/j.pbiomolbio.2016.09.009.

Blasi, D. E. et al.

2016 “Sound–meaning association biases evidenced across thousands of languages”, *Proceedings of the National Academy of Sciences*, ISSN: 0027-8424, DOI: 10.1073/pnas.1605782113.

Bolshoy, A. et al.

2010 *Genome Clustering: From Linguistic Models to Classification of Genetic Texts*, Studies in Computational Intelligence, Springer Berlin Heidelberg, ISBN: 9783642129513.

Bréal, M.

1897 *Essai de sémantique: (science des significations)*, Hachette.

Brunning, J.

1997 “Genuine triads and teridentity”, in *Studies in the Logic of Charles Sanders Peirce*, ed. by N. Van Evra et al., Indiana University Press, pp. 252-270, ISBN: 9780253330208.

Burch, R. W.

1992 “Valental aspects of Peircean algebraic logic”, *Computers and Mathematics with Applications*, 23, 6, pp. 665-677, ISSN: 0898-1221, DOI: 10.1016/0898-1221(92)90128-5.

1997 “Peirce’s reduction thesis”, in *Studies in the Logic of Charles Sanders Peirce*, ed. by N. Van Evra et al., Indiana University Press, pp. 234-251, ISBN: 9780253330208.

Cameron, N. M. et al.

2008 "Epigenetic Programming of Phenotypic Variations in Reproductive Strategies in the Rat Through Maternal Care", *Journal of Neuroendocrinology*, 20, 6 (June 2008), pp. 795-801, DOI: 10.1111/j.1365-2826.2008.01725.x.

Chomsky, N.

1957 *Syntactic Structures*, Mouton classic, Bod Third Party Titles, ISBN: 9783110172799.

1970 "Some observations on the problems of semantic analysis in natural languages", in *Znak, język, kultura*, ed. by A. Greimas and Unesco, Janua linguarum: Series maior, La Haye, pp. 256-260.

Chothia, C. and A. Lesk

1986 "The relation between the divergence of sequence and structure in proteins", *The EMBO journal*, 5 (May 1986), pp. 823-6.

Cole, L. and P. R. Kramer

2016 "Chapter 1.2 - Human Chemistry", in *Human Physiology, Biochemistry and Basic Medicine*, ed. by L. Cole and P. R. Kramer, Academic Press, Boston, pp. 11-16, ISBN: 978-0-12-803699-0, DOI: 10.1016/B978-0-12-803699-0.00016-5.

Consortium, T. E. P.

2012 "The ENCODE Project. An integrated encyclopedia of DNA elements in the human genome", *Nature*, 489 (Sept. 2012), pp. 57-74.

Cooper, J. and D. Hutchinson

1997 *Plato. Complete Works*, Hackett Pub., ISBN: 9780872203495.

Crick, F.

1968 "The origin of the genetic code", *Journal of Molecular Biology*, 38, 3, pp. 367-379, ISSN: 0022-2836, DOI: 10.1016/0022-2836(68)90392-6.

1982 *Life Itself: Its Origin and Nature*, Touchstone Book, Simon and Schuster, ISBN: 9780671255633.

Darwin, C.

1987 *The Origin of Species*, ed. by J. Wallace, Classics of World Literature Series, Wordsworth Editions Limited, ISBN: 9781853267802.

Davidson, E.

2010 *The Regulatory Genome: Gene Regulatory Networks In Development And Evolution*, Elsevier Science, ISBN: 9780080455570.

De Saussure, F.

2011 *Course in General Linguistics*, trans. by W. Baskin, P. Meisel, and H. Saussy, Columbia University Press, ISBN: 9780231527958; orig. pub. as 1916.

De Saussure, F. et al.

2002 *Ecrits de linguistique générale*, Bibliothèque de philosophie, Gallimard, ISBN: 9782070761166.

Deleuze, G.

1988 *The Fold*, Continuum impacts, Bloomsbury Academic, ISBN: 9780826490766.

Deleuze, G. and F. Guattari

1987 *A Thousand Plateaus: Capitalism and Schizophrenia*, Bloomsbury Revelations, Bloomsbury Academic, ISBN: 9781780935379.

Denton, M. J., C. Marshall, and M. Legge

2003 "The Protein Folds as Platonic Forms: New Support for the Pre-Darwinian Conception of Evolution by Natural Law", *Journal of theoretical biology*, 219 (Jan. 2003), pp. 325-42.

Diatka, V. and J. Milička

2017 "The effect of iconicity flash blindness", in *Dimensions of Iconicity*, John Benjamins Publishing Company, DOI: 10.1075/ill.15.01dia.

Dupré, J.

2010 "How to be Naturalistic without being Simplistic in the Study of Human Nature", in *Naturalism and Normativity*, ed. by M. de Caro and D. Macarthur, Columbia University Press, pp. 289-303.

2012 *Processes of Life: Essays in the Philosophy of Biology*, OUP Oxford, ISBN: 9780199691982.

Eco, U.

1984 *Semiotics and the Philosophy of Language*, Advances in semiotics, Indiana University Press, ISBN: 9780253203984.

1990 *I limiti dell'interpretazione*, Milano: Bompiani, ISBN: 9788893440417.

1997 *Kant e l'ornitorinco*, Milano: Bompiani, ISBN: 9788501052148.

2007 *Dall'albero al labirinto: Studi storici sul segno e l'interpretazione*, Delfini Pocket, La nave di Teseo, ISBN: 9788893442275.

Emmeche, C.

2011 "Organism and Body: The Semiotics of Emergent Levels of Life", in *Towards a Semiotic Biology*, Imperial College Press, pp. 91-111, DOI: 10.1142/9781848166882_0005.

Faltýnek, D.

2011 *Sémiotické primitivy v konstrukci gramatik*, Olomouc: Palacký University, ISBN: 978-80-244-3423-0.

2013 "Konstruování gramatik DNA a proteosyntézy ze sémantické báze. [Con-structing Grammars of DNA and Proteosynthesis from the Semantic Basis.]", in *Tygramatika: soubor studií věnovanỳch prof. Janu Kořenskému k 75. narozeninám*, ed. by D. Faltýnek and V. Gvoždiak, Dokorán, pp. 152-184, ISBN: 9788073635442.

Faltýnek, D. and V. Matlach

2016 "Báze nejsou písmena", *SALI*, 7, 1, pp. 20-38, ISSN: 2336-6702.

Fischer, S.

2018 *A History of Language*, Reaktion Books, ISBN: 9781780239460; orig. pub. as 1999.

Foley, W.

1986 *The Papuan Languages of New Guinea*, Cambridge Language Surveys, Cambridge University Press, ISBN: 9780521286213.

Gimona, M.

2006 "Protein linguistics – a grammar for modular protein assembly?", *Nature Reviews Molecular Cell Biology*, 7 (Jan. 2006), 68 EP -.

Harris, R.

2009 *Rationality and the Literate Mind*, Routledge Advances in Communication and Linguistic Theory, Taylor & Francis, ISBN: 9781135838751.

Havlin, S. and al

2003 "Statistical linguistic study of DNA sequences", *Fractals* (Sept. 2003), pp. 269-284.

Henderson, B. and A. Martin

2011 "Bacterial Virulence in the Moonlight: Multitasking Bacterial Moonlighting Proteins Are Virulence Determinants in Infectious Disease", 79, 9, pp. 3476-3491, DOI: 10.1128/iai.00179-11.

Henry, V.

1896 Antinomies linguistiques, sv. 2, F. Alcan.

Hiyama, A., W. Taira, and J. M. Otaki

2012 "Color-Pattern Evolution in Response to Environmental Stress in Butterflies", in *Front. Gene.*

Hjelmslev, L.

1928 *Principes de grammaire générale*, Det Kgl. Dansk. Vidensk. Selsk. Hist.-Filol. Medd. XVI, sv. 16,č. 1, A.F. Høst.

1935 *La catégorie des cas: étude de grammaire générale*, Acta Jutlandica, díl 1, Universitetsforlaget.

1942 "Langue et Parole", *Cahiers Ferdinand de Saussure*, 2, pp. 29-44.

1963 *Prolegomena to a Theory of Language*, University of Wisconsin Press.

1985 "Structure générale des correlations linguistiques", in *Nouveaux essais*, ed. by L. Hjelmslev and F. Rastier, Formes Sémiotiques, Presses Universitaires de France, ISBN: 9782130388272.

Hockett, C. F.

1960 "The Origin of Speech", *Scientific American*, 203, 3, pp. 88-97, ISSN: 00368733, 19467087.

Holm, L. and C. Sanders

1997 "New structure — novel fold?", *Structure*, 5, 2, pp. 165-171, DOI: 10.1016/S0969-2126(97)00176-7.

Huberts, D. H. and I. J. van der Klei

2010 "Moonlighting proteins: An intriguing mode of multitasking", *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1803, 4, pp. 520-525, ISSN: 0167-4889, DOI: 10.1016/j.bbamcr.2010.01.022.

Jacob, F.

1970 *La logique du vivant: une histoire de l'hérédité*, Bibliothèque des sciences humaines, Gallimard, ISBN: 9782070293346.

Jakobson, R.

1932 "Russian and Slavic Grammar: Studies 1931-1981", in ed. by R. Jakobson, L. Waugh, and M. Malle, Janua Linguarum. Series Maior, De Gruyter, chap. The Structure of the Russian Verb, ISBN: 9783110822885.

1971a "Linguistics in Relation to Other Sciences", in *Selected Writings Vol 2: Word and Language*, ed. by R. Jakobson, Mouton, pp. 655-696.

1971b "Signe zéro", in *Selected Writings Vol 2: Word and Language*, ed. by R. Jakobson, Mouton, pp. 211-219.

Jeffery, C.

2005 "Mass Spectrometry and the Search for Moonlighting Proteins", 24 (Nov. 2005), pp. 772-82.

Jeffery, C.

2014 "An introduction to protein moonlighting", *Biochemical Society transactions*, 42 (Dec. 2014), pp. 1679-83.

Ji, S.

1985 "The bhopalator: a molecular model of the living cell based on the concepts of conformons and dissipative structures", *J Theor Biology*, 116, 3, pp. 399-426.

1999 "The linguistics of DNA: words, sentences, grammar, phonetics, and semantics." *Annals of the New York Academy of Sciences*, 870, pp. 411-417.

Karcevskij, S.

1956 "Du dualisme asymétrique du signe linguistique", *Cahiers Ferdinand de Saussure*, 14, pp. 18-24, ISSN: 0068516X.

Karlík, P.

1996 "Syntaktické sloveso", *Naše řeč*, 2, 79, pp. 66-72, ISSN: 0027-8203.

Katz, G.

2008 "The Hypothesis of a Genetic Protolanguage: an Epistemological Investigation", 1 (Apr. 2008), pp. 57-73.

Katz, J. J. and J. A. Fodor

1963 "The Structure of a Semantic Theory", *Language*, 39, 2, pp. 170-210, ISSN: 00978507, 15350665.

Kauffman, S. A.

2000 Oxford University Press, ISBN: 9780199728947.

Keller, H.

1903 *The Story of My Life*, 1930th ed., Om Books International, ISBN: 9789385609398.

Ketner, K. L. and al

2011 "Peirce's NonReduction and Relational Completeness Claims(CP 3.421): in the Context of First-Order Predicate Logic", *KODIKAS/CODE: Ars Semeiotica*, 34 (Interdisciplinary Seminar on Peirce 2011), pp. 3-14.

King, N.

2007 "Amino Acids and the Mitochondria", in *Mitochondria: The Dynamic Organelle*, ed. by S. W. Schaffer and M.-S. Suleiman, Springer New York, New York, NY, pp. 151-166, ISBN: 978-0-387-69945-5, DOI: 10.1007/978-0-387-69945-5_6.

Kister, A.

2015 "Amino acid distribution rules predict protein fold: protein grammar for beta-strand sandwich-like structures", *Biomolecules*, 5, 1 (Jan. 2015), pp. 41-59.

Kobe, B. and A. V. Kajava

2000 "When protein folding is simplified to protein coiling: the continuum of solenoid protein structures", *Trends in Biochemical Sciences*, 25, 10, pp. 509-515, ISSN: 0968-0004, DOI: 10.1016/S0968-0004(00)01667-4.

Kolesov, G. et al.

2007 "Protein knot server: detection of knots in protein structures", *Nucleic Acids Research*, 35, Web Server (May 2007), DOI: 10.1093/nar/gkm312.

Komárek, M. and J. Kořenskỳ

1986 *Mluvnice češtiny 2 : Tvaroslovi*, ed. by J. VESELKOVÁ.

Kořenský, J.

1984 *Konstrukce gramatiky ze sémantické báze*, Studie a práce lingvistické, Academia.

Křen, M. et al.

2015 *SYN2015: representative corpus of written Czech.*

Krivochen, D.

2016 "Divide and...Conquer? On the limits of algorithmic approaches to syntactic structure [Preprint version]", *Czech and Slovak Linguistic Review*, 1(2016) (Mar. 2016), pp. 15-38.

2018 "On Folding and Twisting (and whatknot): towards a topological view of syntax", DOI: 10.13140/RG.2.2.27142.47685.

Krivochen, D. and J. Saddy

2016 "Structure mapping: Uniformity vs. Mixture in Displacement [Preprint]", *Czech and Slovak Linguistic Review*, 2016 (Jan. 2016), pp. 17-46.

Kull, K.

2015 "Evolution, Choice, and Scaffolding: Semiosis is Changing Its Own Building", *Biosemiotics*, 8, pp. 223-234.

Kusebauch, U. et al.

2014 "Mycobacterium tuberculosis supports protein tyrosine phosphorylation", *Proceedings of the National Academy of Sciences*, 111, 25, pp. 9265-9270, ISSN: 0027-8424, DOI: 10.1073/pnas.1323894111.

Lacková, Ľ., V. Matlach, and D. Faltýnek

2017 "Arbitrariness is not enough: towards a functional approach to the genetic code", *Theory in Biosciences*, 136, 3, pp. 187-191, ISSN: 1611-7530, DOI: 10.1007/s12064-017-0246-1.

Langacker, R.

1990 *Concept, Image, and Symbol: The Cognitive Basis of Grammar*, 130./131. Winckelmannsprogramm der Archaologischen Gesellsch, Berlin, ISBN: 9783110125993.

Larson, R. K.

1988 "On the Double Object Construction", *Linguistic Inquiry*, 19, 3, pp. 335-391, ISSN: 00243892, 15309150.

Laubichler, M. D. et al.

2015 "The relativity of biological function", *Theory in Biosciences*, 134, 3-4 (Oct. 2015), pp. 143-147, DOI: 10.1007/s12064-015-0215-5.

Levinthal, C.

1969 "How to fold graciously", in *Mossbauer spectroscopy in biological systems*, ed. by P. DeBrunner, J. Tsibris, and E. Munck, University of Illinois Press, Urbana, IL.

Lévy-Bruhl, L.

1996 *L'âme primitive*, Quadrige (Paris. 1981), PUF; orig. pub. as 1963.

Loose, C. et al.

2006 "A linguistic model for the rational design of antimicrobial peptides", *Nature*, 443, 7113 (Oct. 2006), pp. 867-869, DOI: 10.1038/nature05233.

Mantegna, R. N. et al.

1995 "Systematic analysis of coding and noncoding DNA sequences using methods of statistical linguistics", *Phys. Rev. E*, 52 (3 Sept. 1995), pp. 2939-2950, DOI: 10.1103/PhysRevE.52.2939.

Markoš, A.

2002 *Readers of the Book of Life: Contextualizing Developmental Evolutionary Biology*, Oxford University Press, ISBN: 9780195149487.

2003 "Přírodní zákony a evoluce", in *Filosofie Henri Bergsona: základní aspekty a problémy*, ed. by J. Čapek, Sborníky, slovníky, učební texty, OIKOYMENH, ISBN: 9788072980710.

2008 *Náhoda a nutnost : Jacques Monod v zrcadle nasí doby : sborník statí*, Pavel Mervart, Cervený Kostelec.

2010 "Biosemiotics and the Collision of Modernism With Postmodernity: A Biossemiótica E a Colisão Do Modernismo Com a Pós-Modernidade", *Cognitio*, 11, 1.

2016 *Evoluční tápání: podoby planetárního životopisu*, Amfibios : práce katedry filosofie a dějin přírodních věd Přírodovědecké fakulty Univerzity Karlovy v Praze, Pavel Mervart, ISBN: 9788074652233.

Markoš, A. and F. Cvrčková

2018 "Struktury a artefakty v biologii", in *Language Faculty and Beyond*, pp. 167-162, forthcoming.

Markoš, A. and D. Faltýnek

2010 "Language Metaphors of Life", *Biosemiotics*, 4, 2 (Aug. 2010), pp. 171-200, DOI: 10.1007/s12304-010-9097-6.

Martin, R. and J. Uriagereka

2014 "Chains in Minimalism", in *Language Faculty and Beyond*, John Benjamins Publishing Company, pp. 169-194, DOI: 10.1075/lfab.11.07mar.

Martinet, A.

1957 "Arbitraire linguistique et double articulation", *Cahiers Ferdinand de Saussure*, 15, pp. 105-116, ISSN: 0068516X.

1967 *Éléments de Linguistique Général*, Collection Armand Colin, no. 340. Section de litérature, Librairie Armand Colin; orig. pub. as 1960.

Mathesius, V.

1983 "On the Potentiality of the Phenomena of Language", in *Praguiana*, John Benjamins Publishing Company, p. 3, DOI: 10.1075/llsee.12.05mat.

McClintock, B.

1984 "The significance of responses of the genome to challenge", *Science*, 226, 4676, pp. 792-801, ISSN: 0036-8075, DOI: 10.1126/science.15739260.

Miloš, D.

1957 "K otázce morfologických protikladů (Kritika předpokladu binárních korelací v morfologii češtiny", *Slovo a Slovesnost*, 19, 2, pp. 81-103.

Monod, J.

1972 *Le hasard et la nécessité: essai sur la philosophie naturelle de la biologie moderne*, Points, Essais, Points, ISBN: 9782757844489.

Morris, C.

1971 *Writings on the General Theory of Signs*, Approaches to Semiotics [AS], De Gruyter, ISBN: 9783110810592.

Neuman, Y.

2008 *Reviving the Living, Volume 6: Meaning Making in Living Systems*, Elsevier Science, San Diego, USA, ISBN: 0444530339, 9780444530332.

Paolucci, C.

2004 "Piegature della continuità. Semiotica interpretativa e semiotica generativa", *Versus. Quaderni di studi semiotici*, 97, pp. 111-150.

2006 "Lucien Tesnière autore della logica dei relativi. Su alcune in-sospettate corrispondenze tra Peirce e lo strutturalismo", *E/C*, pp. 1-16.

2010 *Strutturalismo e interpretazione*, Strumenti Bompiani, Bompiani, ISBN: 9788858760208.

Peirce, C. S.

1956 *Chance, Love, and Logic: Philosophical Essays*, G. Braziller.

Peirce, C. S., C. Hartshorne, et al.

1931-1935, 1958 *Collected Papers of Charles Sanders Peirce : Edited by Charles Hartshorne and Paul Weiss*, Collected Papers of Charles Sanders Peirce : Edited by Charles Hartshorne and Paul Weiss, Harvard University Press.

Peirce, C. S. and V. Welby

1977 *Semiotic and Significs: The Correspondence between Charles S. Peirce and Victoria Lady Welby*, English, ed. by C. S. Hardwick, Indiana University Press, Bloomington.

Pigliucci, M.

2009 "An Extended Synthesis for Evolutionary Biology", *Annals of the New York Academy of Sciences*, 1168, 1 (June 2009), pp. 218-228, DOI: 10.1111/j.1749-6632.2009.04578.x.

2010 "Genotype-phenotype mapping and the end of the 'genes as blueprint' metaphor", *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365, 1540 (Jan. 2010), pp. 557-566, DOI: 10.1098/rstb.2009.0241.

Rapaport, W. J.

2007 “How Helen Keller used syntactic semantics to escape from a Chinese Room”, *Minds and Machines*, 16, 4 (Feb. 2007), pp. 381-436, DOI: 10.1007/s11023-007-9054-6.

Robertson, M. P. and G. F. Joyce

2010 “The Origins of the RNA World”, *Cold Spring Harbor Perspectives in Biology*, 4, 5 (Apr. 2010), pp. 3608-3608, DOI: 10.1101/cshperspect.a003608.

Rosenberg, A. and D. McShea

2008 *Philosophy of Biology: A Contemporary Introduction*, Routledge Contemporary, Routledge, ISBN: 9780415315920.

Scherrer, K. and J. Jost

2007 “Gene and genon concept: coding versus regulation”, *Theory in Biosciences*, 126, 2 (Sept. 2007), pp. 65-113, DOI: 10.1007/s12064-007-0012-x.

Schmerling, S. F.

2018 “Rhetorical meaning”, *Linguistic Frontiers*, ISSN: 2544-6339, forthcoming.

Searle, J. R.

1980 “Minds, brains, and programs”, *Behavioral and Brain Sciences*, 3, 03 (Sept. 1980), p. 417, DOI: 10.1017/s0140525x00005756.

Searls, D. B.

2002 “The language of genes”, *Nature*, 420, 6912 (Nov. 2002), pp. 211-217, DOI: 10.1038/nature01255.

Sercarz, E. et al.

2013 *The Semiotics of Cellular Communication in the Immune System*, Nato ASI Subseries H: Springer Berlin Heidelberg, ISBN: 9783642731457.

Sharov, A. A.

2010 “Functional Information: Towards Synthesis of Biosemiotics and Cybernetics”, *Entropy*, 12, 5, pp. 1050-1070, DOI: 10.3390/e12051050.

2016 “Evolution of Natural Agents: Preservation, Advance, and Emergence of Functional Information”, *Biosemiotics*, 9, 1 (Mar. 2016), pp. 103-120, DOI: 10.1007/s12304-015-9250-3.

Sikorska, B. et al.

2012 “Creutzfeldt-Jakob Disease”, in *Advances in Experimental Medicine and Biology*, Springer US, pp. 76-90, DOI: 10.1007/978-1-4614-0653-2_6.

Siltberg-Liberles, J., J. A. Grahnen, and D. A. Liberles

2011 “The Evolution of Protein Structures and Structural Ensembles Under Functional Constraint”, *Genes*, 2, 4 (Oct. 2011), pp. 748-762, DOI: 10.3390/genes2040748.

Skalička, V.

1948 “The Need for a Linguistics of ”la parole””, *Recueil linguistique de Bratislava*, 1, pp. 21-38.

Snyder, M.

2003 “GENOMICS: Defining Genes in the Genomics Era”, *Science*, 300, 5617, pp. 258-260, DOI: 10.1126/science.1084354.

Spencer, A.

1988 “Bracketing Paradoxes and the English Lexicon”, 64, 4, pp. 663-682, DOI: 10.2307/414563.

Stegmann, U. E.

2015 “‘Genetic Coding’ Reconsidered: An Analysis of Actual Usage”, *The British Journal for the Philosophy of Science*, 67, 3 (Mar. 2015), pp. 707-730, DOI: 10.1093/bjps/axv007.

Stevenson, A.

2010 *Oxford Dictionary of English*, Oxford Dictionary of English, OUP Oxford, ISBN: 9780199571123.

Tesnière, L.

1959 *Éléments de syntaxe structurale*, Klincksieck.

Trifonov, E.

1988 "Codes of nucleotide sequences", *Mathematical Biosciences*, 90, 1-2 (July 1988), pp. 507-517, DOI: 10.1016/0025-5564(88)90080-6.

Trifonov, E. et al.

2001 "Distinct Stages of Protein Evolution as Suggested by Protein Sequence Analysis", *Journal of Molecular Evolution*, 53, 4-5 (Oct. 2001), pp. 394-401, DOI: 10.1007/s002390010229.

Trubetzkoy, N. S.

1939 *Grundzüge der Phonologie: Von N. S. Trubetzkoy. Prague, 1939*, Travaux du Cercle linguistique de Prague, Kraus Reprint.

Vachek, J.

1976 "The non-static aspect of the synchronically studied phonological system", in *Selected writings in English and general linguistics*, ed. by J. Vachek, Janua linguarum: Series maior, Academia, ISBN: 9789027930248.

Weaver, I. C. G. et al.

2004 "Epigenetic programming by maternal behavior", *Nature Neuroscience*, 7, 8, pp. 847-854, DOI: 10.1038/nn1276.

Zipf, G.

1949 *Human behavior and the principle of least effort: an introduction to human ecology*, Addison-Wesley Press.