

Ľudmila Lacková: A Linguistic Approach to Protein Folding. Towards a Semiotic Description of Living Systems

Opponent's review of the Ph.D. Thesis

Anton Markoš, Faculty of Science, Charles University Prague

The thesis has an ambition to draw analogy between the content/expression relationship in natural languages, and shape/string of native proteins; briefly, how the linear string of elements (phonemes on one hand, amino acids on the other) constitutes a multi-dimensional “word” that enters the ever-changing dynamics of the world’s “encyclopedia”. The task is the demanding, because – as the author underlines – no “linearity” comparable to the written text (with its digital sequence of letters) is to be attained: we deal with “bodily” entities with many “folds”, “kinks”, etc. I put a value especially to author’s idea of what she calls “participation” – the reciprocal relation (causation) of both domains – instead of more common causal chain from linear to bodily entities in evolution of both language and biochemistry. The speculation is endorsed by her several (single-authored) papers related to the topics. *Undoubtedly the dissertation represents a solid base for future research, and deserves to be accepted as a Ph. D. thesis.*

*

Yet, it is my duty to highlight below in more details both the strengths and drawbacks of both the arrangement of the text, and author’s argumentation. As a non-native speaker of English, I am not competent to evaluate the language of the text; yet I have a strong feeling that reviewing by a native speaker would help it very much; consulting biochemist would help strengthen the argumentation. too. I shall concentrate on (1) the background (starting point), its soundness and terminology; (2) the organization and style of the text; and before all (3) the theoretical solutions and challenges.

1. I am not competent to assess the linguistic terminology; yet what I struggled with is the concept of meaning that appears in many connotations throughout the text. In my understanding, it is the semiotic concept, a relation – as meaning *for* somebody or something: a human being, or, in biosemiotics, for any living being. The subtitle of the treatise suggested such a reading; however, living beings play only a minor role throughout, and much greater attention is paid to semantics, or grammar of relations. I know that biochemists from the Code biology group broadened the concept (as in the expression “polypeptide is a meaning of the genetic code”); it seems that the opinion of the author-linguist on the topic is similar (cf. “p. 23 “word order codes for some meaning”. Meaning as such, or *for* somebody?). More serious is author’s superficial knowledge of established biochemical facts terminology; this makes some statements simply blurred. For example, the author is apparently unaware of the existence of weak interactions that play a crucial role in folding the polypeptide macromolecule: this leads her to take the proper folding of the string for “miraculous”, “obscure” process, a “mystery” (pp. 6, 34), with

“structures of relationships between elements, completely dissociable from their material realization.” (p.7) It is not true even in spoken language, the material realization of vocal cords being decisive in what can be uttered; rules of protein folding may be even more stringent and far from arbitrary (as suggested on p. 25); it is true, however, that endless possible polypeptide strings will “collapse” into only about 1200 shapes in use, the rest being only denatured chunks. Even worse, the first chapter as if suggests that the native protein can attain only one single conformation – as a crystal of a kind. In fact, the very function of *any* protein lies in switching its shape between two or more conformations. The Levinthalian protein is simply an analogue of a “dictionary entry” like infinitive or nominative – even here above such an entry lies a set of morphological states, contexts, clichés, and metaphors. The protein molecule can also acquire a set (sometimes quite big) of posttranslational modifications comparable to diacritics, and changing its shape (i. e., functioning). Under natural conditions it becomes part of multiprotein complexes (“hairballs” counting tens or hundreds of different proteins differently modified and in different positions towards others), loosely connected to other such complexes (“small world” analogy). To give analogical example from Czech language, take the form [sním] connecting vast “hairballs” of dictionary words “sníst/jíst”, “snít”, and “on/ono”. Such homonyms bring about mistakes, dead alleys, but also new, often surprising connections and challenges. It allows also appearance of novelty: the author gives an example of the Czech verb “dát” (give) that is used, as in English, in established grammatical form “sb gives st (in)to”. Recently, however, the Czech verb freed itself from such constraints (“to nedám”). Plethora of similar analogies can be easily transferred to the ecosystem of protein shapes and higher-order structures – up to the global ecosystem.

To raise an interest of cell biologists, the author should first free herself from irritating formulations (example below) that blur the message intended (*italics mine*). The more it is regrettable, that a short consultation with a biochemist or biologist could make the text intelligible. Only some examples:

* “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.” (26) An incomprehensible statement

* “When one sees a peptide chain, they [?] see *nothing other* [as?] an “infinity” of possible structures.” (31) Simply not true, much more can be guessed today.

* “only *protein structure produces amino acids* as discrete units of a particular fold.” (32) Nonsensical statement. Similarly “Also, in the *genetic code*, the *change of an amino acid chain in the DNA* is only a change in substance.” Double nonsense: the genetic code is not the matter of DNA, and DNA is not constituted of amino acids

* For reasons unknown the author uses the term “Modern synthesis” instead of “Central dogma”.

* The notion of “code” is quite fuzzy. The “genetic code” is explained in every textbook, but what is “DNA code”? Apparently the author has regulatory sequences in mind, but is it a code, if it can attain or lose by simple bending the string or binding some protein in the vicinity?

Much more examples could be given here.

2. The composition of the text often gives an air of a collage, with sudden jumps between various topics and authors. Due to the fact, the reading is difficult at places, and I might missed the topics of the author. Apart from this, I consider first, or even first two chapters quite superfluous, going even *against* the argumentation provided in the following chapters. Such an opinion, of course, may sprout from my angle of view; after all, I am a biologist.

3. I see no analogy between protein folding and natural language (as announced in the Introduction), and I didn't get it from my reading the dissertation – but this may be my fault, – I was after more interesting parts. I'd reserve the term “linguistic structure” to sentient beings, i.e. all life – cells, multicellular organisms, communities. The term resonates with my preferred definition of life as a semiotic category. Besides, author's elaboration of Eco's duality of dictionary vs. encyclopedia is valuable in a biological context. Above all however, I praise two points: the idea of participation (participative opposition) of different *domains* (i. e., not levels) on each other, opening a fresh understanding of life's doings – the idea can be, of course, broadened to a plethora of other domains. From participative perspective comes also the last jewel of the work: the idea of reciprocal causation even between domains like protein structure and its amino acid sequence, and more generally, of sublogic relations between life's domains. That leads, first, to speculations – of cell itself, constituting perhaps the decisive part of genetic environment, thus participating on Darwinian selection. It is paramount to develop analogies between life's evolution and evolution of languages and cultures. I urge the author to develop the topics further.

“Life as its own designer” is the title of book from our department (Markoš et al. 2009) this dissertation may give a new turn to the idea. In spite of various mishaps listed above, the learned committee should consider accepting it as a Ph. D. thesis.

Anton Markoš

06 Dec. 2018