PATTERN CHEMISTRY OF LANGUAGE

Yuri Tarnopolsky

The links presented here refer to a series of e-publications PATTERN CHEMISTRY OF LANGUAGE. It is an attempt to apply Pattern Theory of Ulf Grenander (Brown University; see numerous links on the Web) to the atomistic aspects of language. While Pattern Theory is a **mathematical** discipline, Pattern Chemistry is a small contribution to its big domain by a **chemist** with a close connection to Ulf Grenander's work and a lifelong interest in languages, some as different as Russian, Hebrew, Japanese, and Hungarian, apart from main European ones.

The story of this series should better be told from the first person.

I am an organic chemist, Ph.D., who happened to have a lot of spare time to think about matters outside my immediate profession. In the late 1970s, when the political atmosphere in Soviet Russia was growing tenser by the day, I was especially intrigued by the problem of stability of social structures and the further fate of the strained social system. Much later, this direction of thought resulted in "History as Points and Lines," initiated and co-authored by Ulf Grenander. The same question of stability can be asked about the structures of language in the process of historical evolution as well as individual language acquisition.

In 2002, I ran into *The Atoms of Language* by Mark C. Baker. I took his question "what if the words were atoms" completely serious, which was quite natural in the framework of Pattern Theory. That was the initial impetus to the current line of work.

Contrary to a wide spread view, chemistry is not just a description of millions of substances and their transformations. The central organizing idea of chemistry, borrowed from physics, is that between an initial stable state of atoms connected in a particular way, and the final stable state of the same atoms connected in a different way, there is an unstable fleeting state of transition (transition state) from one to the other. The second theoretical idea is that if we have a structure of atoms connected with bonds in a certain order, then any other arrangement of the same atoms can emerge spontaneously. In fact, chemical reactions result in only a few products because the rest of transformations are negligibly slow. The third idea is that the speed of transformation depends on the height of the transition barrier between the initial and final state. Pattern Theory provides a measure of the stability of abstract structures of any origin, including those of thought and its expression.

Obviously, these ideas can be easily generalized over other objects built of atomic entities and connections between them, such as structures of language. Pattern Theory from the point of view of a chemist is exactly this kind of generalized chemistry. If applied to

language, Pattern Chemistry regards its generation and evolution as a natural process governed by relative stability and kinetics of linguistic structures.

Pattern Theory of Ulf Grenander studies structural complexity regardless of interdisciplinary boundaries. It reduces structure to a set of atomic entities (generators) selectively connected by bonds, thereby representing observable objects of widest variety, including language and thought. The unusual aspect of Pattern Theory is its metrics which allows for distinguishing between more and less stable (i.e., probable) structures. **Pattern Chemistry** focuses not so much on stable structures as on the fleeting transition states between them, similarly to the way chemistry treats molecular transformations, making distinction between fast and slow transformations.

The parallel between linguistics and chemistry has been a standard explanatory model since the discovery of the DNA's structure and its ability to carry a protein "meaning." Besides, chemistry uses a particular language (chemical nomenclature) to convert complex non-linear structures ("chemical thoughts") into a linear word which can potentially be communicated through speech.

The central ideal of Pattern Chemistry is that **complexity** in nature and society evolves **from simple** states **by simple steps**. This perspective of complexity differs from the concept of dynamic complexity represented by systems of differential equations in the so-called "order-from-chaos" theories. Indeed, this is an "order-from-order" concept, more in line with the Darwinian paradigm of evolution. It seems interesting to investigate language acquisition, generation, and evolution in the discrete stepwise way, which is very similar to both ontogenesis/phylogenesis in biology and chemical transformations.

Thus, e-papers SALT and SALT 2 explore the stepwise extraction of grammar from the text of the Hungarian folk tale "Salt," basing on extremely simple rules. I used the Hungarian text in order to eliminate all semantic associations natural for a native speaker of English and to emphasize the universality of the model regardless of grammar and vocabulary.

The most recent e-publication in this collection, entitled, maybe too ambitiously, PATTERN CHEMISTRY OF THOUGHT AND SPEECH, comes closer to the evolutionary problems that any of the preceding ones.

Human thought is not directly observable, apart from vague shadows in brain scans. We simply do not know what it is. We have nothing to support the belief that we think in linear sequences of whatever units, but we definitely speak in them. Moreover, the syntactic trees are definitely non-linear. The paper further explores a hypothetical protolanguage, called Nean, in which the simple elementary thought consisting of two connected entities directly translates into the simplest elementary phrase consisting of two words. Nean, unlike extant languages, is topologically identical with thought.

Nean sounds like a repetitive random series of elementary doublets. Two doublets with common element can be combined into linear triplets. The e-paper explores the ability of this inherently linear and primitive language to express more complex non-linear thoughts by means of the process of linearization. It appears that Nean, subjectively, is quite expressive, in spite of its primitiveness.

I present a computer simulation, based on simple principles, which represents thinking and speech as competition of alternative thought structures for a spot in consciousness and further generation of linear speech-ready expressions longer than elementary doublets. I use the story of the Three Little Pigs as the substrate of the process and discuss the potential of Nean for further complexification and grammaticalization.

I regard Nean as a point of evolutionary divergence between thought and speech, both initially linear, after which a variety of non-linear grammars can emerge. The current collection of e-papers, or, rather, essays, illustrates the way to this idea, starting from MOLECULES AND THOUGHTS.

Kinetics is the study of speed. In short, the main idea is: we say what we can say **faster** and we say **faster** what we can **faster** convert from nonlinear thought into linear speech.

Unlike distant worlds and invisible atoms, language is as directly observable, rich, and exciting as the beauty of nature. Not accidentally so many outsiders felt the irresistible pull of linguistics and tried their mind in solving its puzzles. I believe that the current collection, coming from remote outskirts of liguistics, can illuminate the problems of evolutionary linguistics and intimate mechanisms of verbal communication from a kinetic angle and in chemical light.

> **NOTE**: The kinetic direction has already been tested in linguistics by Martin Nowak *et al* who invoked the fundamental idea of competitive kinetic evolution (Manfred Eigen).

LINKS

1. **Molecules and Thoughts: Pattern Complexity and Evolution in Chemical Systems and the Mind** , 2003 (or this and this).

2. **TIKKI TIKKI TEMBO and the Chemistry of Protolanguage** , 2004 (or this).

3. **Pattern Theory and "Poverty of Stimulus" Argument in Linguistics** ,2004 (or this) .

4. **The Three Little Pigs : Chemistry of language acquisition** ,2005 (or this).

5. **Salt: The Incremental Chemistry of Language Acquisition** , 2005 (or this).

6. **Salt 2: Incremental Extraction of Grammar by Simplistic Rules** , 2005 (or this).

7. **The Chemistry of Semantics** , 2005 (or this).

8. **Do Pirahã speak Nean?** , 2007 (or this) .

9. **PATTERN CHEMISTRY OF THOUGHT AND SPEECH** , 2010 (also here) .

9A . **APPENDIX to PATTERN CHEMISTRY OF THOUGHT AND SPEECH:** MATLAB codes , (also here) .

EMAIL: SIMP@SPIROSPERO.NET

MAIN PAGE: http://spirospero.net/complexity.htm

HOME PAGE: http://spirospero.net

November, 2010