THE CHEMISTRY OF SEMANTICS

Yuri Tarnopolsky

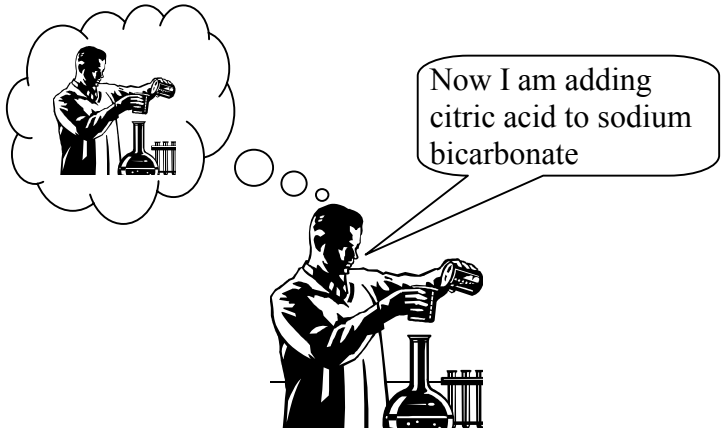
2005

ABSTRACT

This preliminary e-paper continues the examination of language as a quasi-molecular system from the point of view of a chemist who happens to ask, "What if the words were atoms?" The general principles of atomism, locality, simplicity, and transition state are applied to the unobservable structure of thought, regarded as a configuration in Pattern Theory (Ulf Grenander). The non-linear thought interconnects maximum three atomic observables and the coordinate of change in the configuration. The thought is subject to linearization in order to be translated into an utterance. The verb is regarded as the lexical marker of the coordinate of change.

KEYWORDS: semantics, thought, idea, meaning, language acquisition, protolanguage, Navajo, language, pattern theory, chemistry, chemical nomenclature, speech generation, Ulf Grenander, Leonard Talmy.

What if the words were atoms?



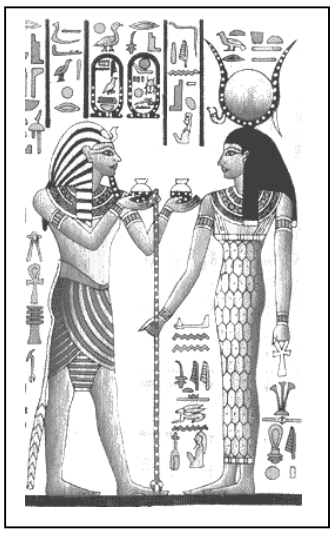
THE CHEMISTRY OF SEMANTICS

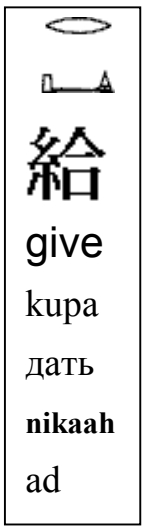
Yuri Tarnopolsky

2005

1. EXAMPLE

Left side: Horemheb gives drinks to Hathor.
A copy of a picture from the tomb of Horemheb, who became pharaoh around 1321 BC.
Right side: Written representation of the idea of giving in Egyptian, Chinese, English, Swahili, Russian, Navajo, and Hungarian.
I see **the pharaoh**, **the goddess**, and **the drinks**. But where is **the give**? And what about **the to**?





給
give
kupa
дать
nikaah
ad

2. INTRODUCTION

It is quite remarkable that the subject of meaning, with which everybody is intimately familiar, is so evasive. The mind is still an open frontier for thought.

The literature on semantics and its various aspects is enormous [1] but not really contentious. It seems that any contributor to it is able to offer his or her own system and notation to represent meaning, without joining an army, however.

As I believe, the lack of a consensus combined with the absence of an intense debate is a good evidence that the subject of a discussion either does not exist or should be looked at from a different angle.

There could be little, if any, debate about the meaning of the absolute majority of particular words (although even the existence of word is debated), but not about **meaning** itself. Does meaning exist? Is it anything else than what we find in any monolingual dictionary? For my part, I am completely satisfied by the article **meaning** in Webster's II:

1. Something signified by a word. 2. Something one wishes to convey, esp. by language. *Etc*.

If I am content with the obviously vague "something," it is because I usually know what **I** myself mean (but see my concluding remarks). The problem arises when we realize that although the dictionary definition is the result of a consensus, no dictionary can tell you what Mr. Smith means right now by saying "I don't like it at all" or Mrs. Johnson wished to convey yesterday by saying "So nice to see you." Are **meaning** and **intent** the same? There are split opinions, too. One can find in the literature a collection of mental variations on the theme of meaning as Baroque as Bach's *Goldberg Variations*.

> It is, in short, music which observes neither end nor beginning, music with neither real climax nor real resolution, music which, like Beaudelaires's lovers, "rests lightly on the wings of the unchecked wind." Glenn Gould http://www.rjgeib.com/music/Top-Ten/gould.html

I am neither a linguist nor a philosopher. All I want is to launch the problem on the wings of the chemical wind and see whether it flies.

As somebody who did some volume of translation, I can testify that there is no general problem in rendering meaning across the language barrier. All problems of translation, however pervasive, are minor, specific and local. Although there is no way to know what exactly people meant by writing a sentence in Hungarian or Japanese, we can usually have a very good guess and normally know it precisely. Even machine translation can do a good job in most cases. Moreover, machine translation within a very narrow professional context, for example, weather or medicine, where consensus dominates, can

be very successful. The absolute majority of professional languages are tailored to avoid (and rarely to enhance) ambiguity. All I needed in order to translate a piece of technical literature from a foreign language was to understand how the described device or system looked, worked, or behaved. I did not care about what the author thought. To understand meant to say it in my native language, although some moment of a sudden revelation could precede it in hard cases.

I believe that theoretical semantics is an attempt to represent by means of human language what by definition is as little observable as a deity: human **thought**. This apparent paradox dissolves if we separate language from thought and thought from meaning. Language **is** the language of thought. Period. Without language the thought is mute. But what **is** thought? Language is a kind of order. Is thought a kind of disorder? This verbal prestidigitation makes some sense to a chemist.

This e-paper continues the examination of language as a quasi-molecular system from the point of view of a chemist who happens to ask, "What if the words were atoms?" [2]. By analyzing the text of a folk tale as a sequence of short fragments and following a few simple rules we can see how the language acquisition creates a quasi-chemical system in which some sequences ("molecules") are more probable than others and some are quite improbable. For example, in the system of the *Tale of The Three Little Pigs* [2D] we can generate sentences **Pig built a house** and **Man gave the straw**. Although **Pig gave the straw** and **Man built a house** (as well as other combinations) are not supported by the context, the pattern **Subject Verb Object** is. This all is well known in computational linguistics and all a chemist could add is a different mode of thinking, which I call—defiantly—**simplicity**. It is natural for a chemist to regard anything complex as a structure (molecule) of atomic (i.e., **simple**) objects held together or rearranged by local (i.e., **simple**) interactions.

> Although the idea of a computing homunculus is deeply alien to chemistry, the idea of chemistry is not alien to computation [2G]. Neither is the idea of language to chemistry.

The problem, illustrated by the above examples, from pharaoh to pig, is: what happens right before we irreversibly release the utterance into the air and even before we silently formulate it to be released ?

3. CHEMISTRY AND CONFUSION

Chemistry found itself in the state of a teenage confusion right before turning into science about one and a half century ago. It was clear that matter consisted of some unobservable units that could be combined in different quantities. This paradigm, as old as science itself, had been laid out in eloquent verses by Lucretius, the Roman devotee of much less preserved by time Democritus the Greek.

What chemistry accomplished, due to Alexander Butlerov (1828-1886) and Friedrich August Kekule (1829-1896), was to separate the chemical reality, as it is seen—but not understood—through the transparent wall of the flask, from what we cannot see but can understand. It was done by a kind of minimalist program: we have no idea what is going on there, but we are able to deduce which Democritean entities (atoms of chemical elements) are in the substance and which of them are connected.

Starting with Jacobus Van't Hoff (1852-1911), the very first Nobel Laureate in chemistry (1901), chemistry has been eagerly acquiring from physics all the new gadgets invented to see the atomic details directly. Van't Hoff was among the first to look molecules right into the face (and what he saw was a kind of the double image which the drunks are believed to see). Before the modern instrumentation, the fact of the absence or presence of a chemical bond between two atoms could be only deduced logically from the observable behavior of the molecule, for example, by chemically splitting **A—B—C—D** in the middle and identifying the

fragments **A—B** and **C—D**. In the art of connecting the dots the chemists had to have the skills and talents of Sherlock Holmes and Dr. Watson combined.

If thought is represented as atomic entities connected in a certain order, which is what semantics does for living, the way chemistry handled the invisible may be instructive for semantics. In what order? This we can deduce from human behavior, especially, from speech. But, as chemistry discovered at some point around 1950, the real state of molecular affairs in the flask was a kind of creative chaos that could not even be denoted by common chemical formulas. This discovery of the role of the fuzzy and irregular transition state in chemical transformation was the final act of maturity for chemistry as science—which is not the same as ultimate wisdom, of course. What we can learn from chemistry is how the evasive and fleeting state of confusion (thought) creates certainty (language).

> Should we wait until neurophysiology provides us with some kind of a device to observe the thought? This is a terrifying prospect for a human being who will turn into a simple screw of the social machine, so that the pipe dreams of Joseph Stalin would become reality overnight. Fortunately, there is a big snag for the physiology: the brains of the poor mice are split open to researchers, but mice do not think the way humans do, if they think at all. The brains of the humans—the last wilderness on earth waiting for a bulldozer—are for a while relatively protected, unless spilled out electronically.

There are two areas of chemical reality. The first one includes the stable terminal objects such as molecules (atoms are not always stable). The language for verbal one-to-one description of molecules of any size and structure is known as chemical nomenclature [3]. It can be found in any textbook of organic chemistry. In principle, it is similar to a system of rules and conventions used to identify the appearance of an animal in zoology, a plant in botany, or a human appearance in criminology, but with a substantial difference: the atomic composition and the connections between atoms can be completely extracted from the linear verbal description and *vice versa* because of a one-to-one correspondence between them, although molecular structure is typically non-linear. As I repeated many times [2], this unnoticed elephant in the room is well worth attention of linguistics and semantics

The second area comprises the chemical change, i.e., variations and transformations of chemical composition and structure. The chemical names are all nouns, but the terms of chemical dynamics are mostly either verbs or related verb-derived nouns, (like **to substitute** or **the substitution**).

There is another subtle but important aspect of Chemicalese. If a chemical paper (and for that matter any scientific paper) contains a description of an experiment, it is done in a particular, significantly standardized language which, like the language of chemical nomenclature, is designed to ensure the exact reproduction of the experiment by any other chemist, as the following fragment [4] illustrates:

> **Exp. 4.** *Reaction of Naphthazarin (**3**) with Cyclohexa-1,3-diene*
> To a solution of naphthazarin (**3**) (1.9 g, 10 mmol) in toluene (40 cm^3) was added cyclohexa-1,3-diene (2.43 g, 30 mmol). The mixture was heated at reflux under argon for 3 days. After removal of the solvent and residual diene by evaporation, the crude cycloadduct (2.62 g) was obtained.
>
> Dr. Jonathan Miller's home page, from which I quote, is a delightful induction into the atmosphere of organic chemistry.

An especially concise style of experiment description is used in *Chemical Abstracts*. This kind of language is intentionally primitive as compared with the more sophisticated (although dry enough) language of chemical theory. The latter, however, loses all its luster against extremes of Legalese and Greenspanese at the high end of the art of obfuscation. Anyway, the rest of the chemical paper—introduction, review of literature, discussion, and conclusion—is not as rigid as the experimental part.

As a stalker of the invisible, I limit my linguistic interests to human language in its earlier stages, when it was emerging from the hypothetical protolanguage Nean [2B] as a no-frills means of communicating information in a live coverage style. We can see a reflection of this stage in the development of language by infants and little children.

By definition, information is conveyed if a message is expected with sufficiently low probability, such as about appearance of a predator, prey, or a loss of a hunter. The language of chemical experiment is a good model of the early evolutionary stage when the things are conveyed typically only if there are alternatives. I see the upper end of this particular stage as the emergence of plans, desires, myths, and tales describing **imaginary** or **unobservable** events in a maximally simplified (from modern point of view) matter-of-fact language.

> It is rarely possible to translate good poetry. Nevertheless, Ivan Bunin's famous translation of Henry Longfellow's *Hiawatha* into Russian was regarded by some as better than the original. I explain its success by the descriptive, explicit, and universally human nature of the mythological material.

Following the example of chemistry, let us call the atoms of thought **ideas** and acknowledge that they can be of two kinds (innumerable literature has been piling up since times immemorial). One kind comprises the terminal ideas that stand for (mean, denote, signify, relate to) individual objects: **dog, pine, man, shadow**, at which we can point a finger. Qualities and quantities, like **red, big, slow**, form bonds with them but are not material in the same sense **dog Spot** is. Neither are verbs, like **run, sit, eat**. They all belong to the second kind, together with other ideas like **from**, **inside, under**. I am not ready to suggest any terminology except **primary** and **applied** ideas. **The give** and **the to** in the Example with which this paper starts are applied ideas, invisible *per se*. Thought is a set of ideas. A thought itself can be an idea in another, more complex thought.

4. THE GIVE AS AN IDEA

The give, which I was looking for in the Egyptian wall painting, is something that I do not know what it is *per se*. It occupies a place in my mental construct involving other components, which unlike **the give** are directly observable and belong to the first kind. It has some relation to the verb **give**. No wonder I cannot see **the give** because it is an applied idea—as applied as the paint to the wall of Horemheb's tomb. We need some substrate to apply it to.

The following is a tentative imitation of a chemist's approach to thought as a molecule. This parallels the approach of Pattern Theory [5] in which atoms are called **generators** and molecules are **configurations**. In addition to that, Pattern Theory systematizes configurations and their transformations as patterns, which chemistry also does.

Imagine—departing from Egyptian paintings—that we are watching a real life or soap opera episode with **Ken (K)** and **Lucy (L)**, two people we can see and represent as primary ideas. **Ken** is **giv**ing (**g** , applied) **Money** (**M**: a bag of collectible quarters, also a primary) to **Lucy** right before our eyes, from hand to hand.

This is how the episode could be described in Nean [1B]:

> **Ken. Lucy. Ken Money. Ken give. Ken Money. give Money.**
> **give Lucy. give Money. Money Lucy. Money Lucy. Lucy Money.**

Do not underestimate the raw expressive power of Nean, rich enough to do even without the verb **give**:

> **Ken Money. Ken Money. Ken Lucy. Ken. Lucy Money. Lucy Money.**

> Something like Nean is used by sports commentators, especially for the fast moving soccer ball passed from a player to player: "Pele. Didi. Pele again. Didi. Pele."

Used in such manner Nean has no problem with differentiating between the donor and the acceptor if the narrator and the listener speak the same dialect of Nean. If Lucy was the donor, we would start with **Lucy Money** and end with **Ken Money**. The natural temporal flow of speech in the Nean grammar is the verb tense marker even without the verb.

From my perch, Nean represents the very moment of conception of the verb and tense categories, as the old chemists used to say, *in statu nascendi*. A linguist could say that **Money** performs both noun and verb functions, quite natural for English, as the word **hand** does. So to speak, **Ken monies Lucy**. Compare with hypothetical **Ken hands his hand to Lucy** or **Ken clubs a lion with a club**. While the first is ambiguous (touches? offers? proposes?), the second is pragmatically unambiguous even as **Ken club lion** or **lion club Ke**n.

There could be a different solution for differentiating between **Ken kill lion** and **lion kill Ken**, as well as **Ken break bone** and **Bone break Ken**: to add semantic markers of distinction between a human and an animal, as well as animate and inanimate, or even between long and short, flat and round, etc. The nominal classes of Swahili and the hierarchy of nouns and the classificatory verb stems of Navajo are examples. The evidence of the original fusion of syntax and semantics is also seen in the noun classifiers of Chinese ("measure words") and counting suffixes of Japanese.

Semantic markers seem natural if the language evolves from overgeneralization (which happens with infants, too) when, for example, not only all liquids, but also some far-reaching chains of associations can have the same sign.

In Sumerian, for example [6]:

a, e$_4$: n., water; watercourse, canal; seminal fluid; offspring; father; tears; flood.

ú: n., plant; vegetable; grass; food; bread; pasture; load

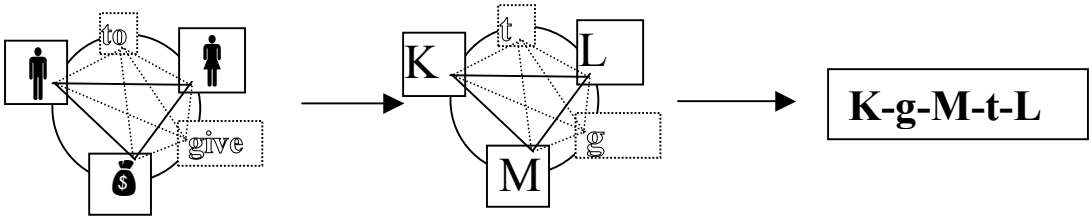
See an interesting discussion of this subject in [6].

To speculate further, the relatively redundant agreement, as between nouns and adjectives in Russian and Swahili, may be an artifact of Nean. I would even dare to suggest that all the markers come from the primary ideas of the inherently redundant Nean. They probably evolved as means of speeding up the transfer of information, which was, I think, the main driving force of early language evolution, quite similarly to the evolution of species, in modern terms, and evolution of technology in modern times.

Next I will try to develop a chemistry-friendly representation of the non-linear thought that generates the linear utterance **K-g-M-t-L**, in which the letters are atomic symbols for words.

I regard the pre-historic thought as a simple **aggregate** of a few ideas that are **all interconnected** just because they are together by some criterion, namely, belonging to the same thought reflecting the same spatial or temporal closeness. One could ponder whether the topology of thought is a full graph or the zero topology of a (simple) set in which elements are labeled, but neither connected nor have duplicates.

Set with duplicates is called bag. Set is what collectors display in a collection where items do not have duplicates or as words in a dictionary. One can also remark that if thought is just a set, animals think too, which I would not dispute. Humans, however, went much farther, but how far—we will never measure until we know the starting point.

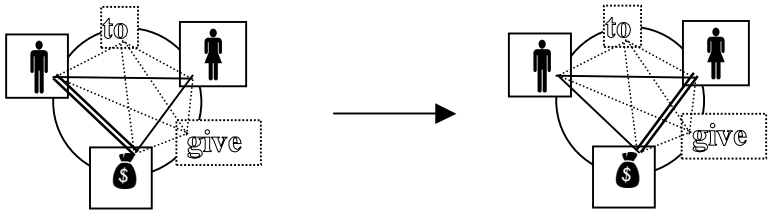


Perception **Thought** **Speech**

(1) (2) (3)

We can point to the primary and stable **K**, **L,** and **M**, but not to **give** or **to**, unless in speech. Let us tentatively reserve some space for the latter two in the perception and thought, denoting their ghostly presence with outline font and their connections with dotted lines. The lines mean that **all** the atoms of thought are **more or less** connected in the object of thought, although we may not know how exactly. This point is what distinguishes the approach of Pattern Theory [5] and chemistry to structures. The structure, whether molecule, thought, or utterance, is a **configuration,** while pattern is, approximately, a class of configurations, for which the rules of similarity within the class are formulated.

In order to notice the action, we need to look at the event at least twice:



(4) (5)

Having done that we can, at last, notice that there are some changing connections between the ideas, invisible in a static picture. The strong (double line) bond between **M** and **K** (4) moves to the position between **M** and **L** (5). To put it differently, **M** moves (migrates) from **K** to **L**.

In a chemical style notation:

$$K—M + L \rightarrow K + M—L \quad (6) \quad \text{or} \quad \begin{matrix} K—L \\ | \\ M \end{matrix} \rightarrow \begin{matrix} K—L \\ | \\ M \end{matrix} \quad (7)$$

This transformation is a pattern, applicable to many primary ideas, and pattern is an **overgeneralization**, much exceeding in its range the Sumerian notion a (**water,** etc.).

Focusing on the points of change, we come to the "chemical" equation, in which the broken lines symbolize the transition state:

$$\mathbf{K—M + L \rightarrow [\ K\cdots M\cdots L\ ] \rightarrow K + M—L} \qquad (8)$$

Chemical transformations in an isolated system are reversible and the transition state can lead to either the initial or the final state. The uncertainty can be reflected in language: **Ken is considering giving money to Lucy.** Or even: **K and L fight for M**. For an irreversible action, the single arrow indicates the direction of the transformation.

Regarding transformations, the principle of atomism means that they can be decomposed into simple and further indivisible steps. In the spirit of chemistry we can define **simple** (atomic) acts as occurring within a topological neighborhood. The **complex** acts can be reduced to a sequence of simple ones. This principle strongly limits the variety of simple actions:

A—B → A B	Ken Money → Ken, Money	(9)
A B → A—B	Lucy, Money → Lucy Money	(10)
A → Ø	Money → (no money)	(11)
A Ø → A B	Adam → Adam, Eve	(12)

Even already simple enough transformations can be represented as a sequence of the above simple steps:

A—B → A—C Sour grape → sweet grape

A—B → A Ø ; A C → A—C (13)

A—B—C → A—B C Dog bite man → Man bite dog

A—B—C → A—B C → A B C → A C—B → C—B—A (14)

There are **maximum four** atomic components of a change: formation (1) and breakup (2) of a single bond and appearance (3) and disappearance (4) of a generator. Jumping to thought, a simple thought **P** (the stem cell of all thoughts) is, at most, a quartet:

$$\mathbf{P} = \{\ \mathbf{A, B, A—B\ ,} \rightarrow\ \}\ , \qquad (15)$$

where **A** and **B** are ideas, **A—B** is a bond between them, and arrow → is a coordinate of change, i.e., the locus of change, for example, appearance/disappearance of an atomic idea or locking/breakup of the bond. The arrow, for which **pointer** is a better term, **points** to the changing component by connecting the initial and the final states of an act of change. There are at least two possible notations for a change (16, 17):

A—B → A Ken Lucy → Ken (alone) (16)

A—B
  ↑

A—B → A—C Sour grape → sweet grape (17)

A—B , A C
  ↑ ↑

The arrow (pointer) is, in my opinion, the precursor of the verb. In early steps of language evolution it was probably a single sound or a gesture signifying or symbolizing action (the theories of gestural origin of language origin are vigorous today). It would be used only if the context was obscure.

A—B
 \ /
  C
(18)

Note that the "atomic" configuration **A B C** in **(14)** makes no sense in linear speech where the adjacency is already a bond. Neither does a cyclic configuration **(18)** , equally impossible in the linear speech. Both can be seen as short-living unstable stages of the process of linearization of thought into speech . What is wrong as speech is OK as thought.

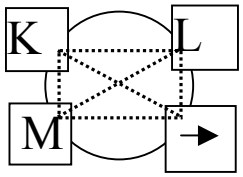
Transition state in chemistry is anything but state. It is a **process** which can be approximated by a sequence of elementary steps, not in all-or-nothing fashion, but as a continuously changing distribution of bonds of variable strength. I must emphasize that the chemical transition state is still as little observable as the thought, although the situation could change. What helps, is that the elementary steps are always **local.**

5. FROM THE POINTER TO THE VERB

Strictly speaking, **the give** (the thought, not the verb) consists of four elements:

{K, M, L, → }

I hypothesize that the thought **give** is a transition configuration between the representation of the perceived scene in the mind (which is outside linguistics) and the utterance that conveys the thought.



The give

Ken give Money to Lucy

or

Ken give Lucy Money

(19)

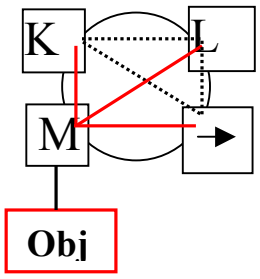
We still cannot put a finger on the arrow, but whatever it is, the four-component thought transition state is highly ambiguous: it can be linearized into utterance in 4!=24 ways. The alternative form **Ken give Lucy Money** is unambiguous only because **Lucy** is animate and **Money** is not, which we know, but, unlike in Swahili, do not mark. **Ken put**

box bag is highly ambiguous because box and bag can be put into each other. In a language similar to Navajo and more elaborate than Swahili, the difference can be marked by verbal object suffixes, and it in fact is marked in Navajo [7], where bag and box require different roots for verb **give**: ni'ą for give box and niyį́ for give bag. The ambiguity can be eliminated if the position in the four-word sequence alone signifies the semantic role, which even the morphologically skimpy English language does relatively rarely.

Grammar, from the chemist's point of view, is a way (more accurately, a **catalyst**) to reduce the stress of the transition state. In terms of chemistry, it gives a preference for linearizing the thought in a particular way, reducing the entropy of choice, although the grammar can often be violated without detriment to understanding. The post-Nean grammar works by either attaching various markers or freezing the word order, or both.

While the word order has its limits (only not more than three ideas form a linear neighborhood A—**B**—C), even a single morphological marker can help a lot.

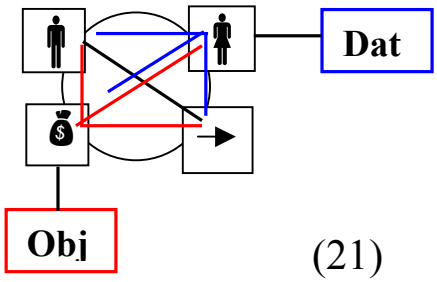
If we attach a marker (in speech, not in thought!) of the direct object (**Obj**) to **Money**, for example, all three its transition bonds (red) are not needed anymore:



(20)

Money, with its marker (infix or case ending), can be now placed anywhere in the utterance, although a certain predominant word order characterizes all languages. This ordering is necessary because the remaining transition triangle **K, L, g** is still too ambiguous to linearize.

If we add a marker for the indirect object (**L**) , in the form of a dative case (**21**) or a preposition (**to**) , the only remaining ambiguity is the relative positions of **K** and **g**.



(21)

It can be disambiguated (I hate this word!) in two ways: by marking either the subject or the pointer.

Japanese has a prominent marker for the subject (**-ga**) , as well as for both objects (**-o** and **–ni**) , plus a topic marker (**-wa**). Nevertheless, the end position of the verb, which is also often marked with a suffix (**-mas** , –**suru , -iru,** etc.) is fixed. Hungarian, which marks the objects (with **-t** and –**nek**) but not the subject, has no fixed word order, probably because the verb, **lexically different from noun**, is almost always marked. The Japanese sentence looks like a single word with fixed positions for all components, clearly marked. Swahili resolves the ambiguity by neatly packing all semantic object markers into the verb, already loaded with tense, but it still needs an **SVO** word order to keep both **S** and **O** close to **V**. Russian typically marks all it can mark but the subject, so that the unmarkedness of the subject is the marker itself, and Russian, like Hungarian, has a very free word order. Navajo lists the nouns in the very beginning in a certain order, according to their rank of animation:

> *Human → Infant/Big Animal → Medium-size Animal → Small Animal → Natural Force → Abstraction*

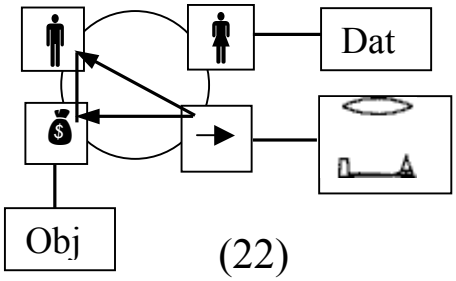
Such differences between the languages could be a fossil pit for paleontology of language.

How can we mark an arrow which, theoretically, is always the same and should be assigned a single word? Whatever marker we attach to the same unique word, it will be

just another unique word, which, by the way, does not need any marking. The **verb** is just the lexical marker of the pointer, which is not the only way to do it because **the verb**, strictly speaking, **is not necessary**. Parts of speech, syllables, words—are they a matter of convention? Certainly not in English. But together with some linguists I strongly suspect that modern linguistics as the science of language bends, trembles, and groans under the heavy Indo-European burden imposed by Panini's grammar of Sanskrit. The ancient speaker of the emerging language never knew what terms the grammarians would use afterwards.

> As a free-thinker, I feel a great deal of mistrust when the linguists speak about **SOV** word order when the verb is **always** at the end. This is the **verb-last** order, people!

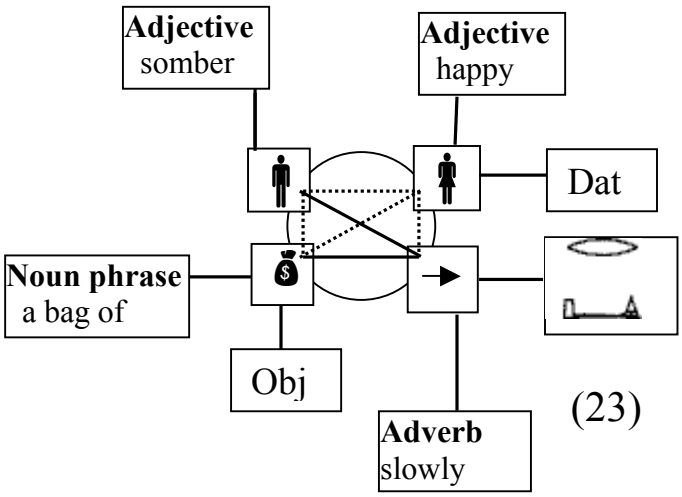
Language evolved as patterns of thought and patterns of its linearization, which, as we see, can be done in a vast variety of ways. Neither word nor verb is a universal concept of language. Furthermore, the patters of thought are representations of patterns of the world—the idea clearly expressed by Spinoza—but all we know about the world is our representation of it. The thought in the following interlingual form **(22)** is, probably, sufficient to translate it into any language equilibrated with a relatively simple and stable reality. The arrow here points to **K** and **M**, which means that Ken has to part with his Money and Lucy is ready to accept the bag (we don't know if she will) but not necessarily Ken himself. The Egyptian hieroglyph ***rdi*** is chosen here to represent the idea of **give.**



(22)

Why then did the verbs develop at all if they were not necessary? This is something paleolinguistics should be concerned with. It is a pure speculation on my part, but I suggest that the patterns of human thought arose from the use of tools, which is rather atypical of animals. To kill, eat, mate, share, cheat, and fight required two components of thought. The tools and the hands to use them in various ways were, probably, responsible

for the development of three- and four-component thoughts. Much more can be speculated on this platform, but I better stop.



(23)

Structure (**23**) gives some hints how thought could further develop complexity, while preserving its ancient core and compatibility with linearization, but it is beyond the scope of my strictly provisionary paper.

5. NOTE ON **GIVE** IN NAVAJO (in which I am by no means an expert).

Polysynthetic (a chemistry-smelling term!) languages, like Navajo, linearize thought in a very strange for Europeans, but basically natural way: the markers are attached not to the nouns, but incorporated into the "verb" and show who or what does what to whom or to what in what manner. In other words, the "verb" is the skeletal phrase, a linearized pattern of thought, preceded by the unmarked nouns [7]:

Łéechąą'í mósí yinoołchéél
dog cat it-is-chasing-it
the dog is chasing the cat

Abandoning the firm chemical ground under my feet, I would say that the above example looks to me as the embryonic "stem cell" for all languages.

The structure of Navajo seems overwhelmingly complicated, primarily because of the prohibiting complexity of the available printed dictionaries, but with some doggedness a chemist can see how it works. The language of chemical nomenclature [3] is a close relative of polysynthetic languages. Example [4]: 1,4,4a,9a-*tetrahydro*-5,8-*dihydroxy*-1,4-*ethano*-9,10-*anthraquinone*. See [4] (Exp. 4) for what it depicts.

It is my uneducated guess that the elaborate polysynthetic languages, like Navajo (not too many of them are well studied), do not have verbs in Indo-European sense. Instead, they use **ideograms**, similar to Egyptian and Chinese ones, as well as to the German compound words, built not from graphic elements but from phonemes. The pictorial nature of Navajo ("mental television") was noted by the native speakers.

> Why are tribal languages often very elaborate and sophisticated? Because thousands of years of stable ordered life with little social and political turmoil are beneficial for the slow evolution of complex patterns of culture. Just compare with the dumbed down, primitive, fragmented, kaleidoscopic, superficial, ephemeral, hectic, and hysterical culture of the twenty-second century.

The Navajo "verb" is a whole proposition [7A,B] , which means that it is actually the entire phrase in which the primary components (Subject, Direct Object, Indirect Object) are always the same standardized elements, similar to pronouns and, if you wish, to the suffixes of classes in chemical nomenclature: ***hydroxy*** or ***ol*** for alcohol, ***di*** for dual plural***, on*** for >C==O group, *etc*. The primary elements are listed in the very beginning, usually without markers. By analogy with Japanese, where the topic of the proposition is a ***wa***-ending noun or noun group in the very beginning, I would call Navajo a **polytopical** language: all its nouns are the participants of the action/state, and not just one of them, as in Japanese. They are like listed roles and actors in titles with which older

movies begin—but the stars go first—giving the clue to the components of the subsequent "verb."

One can imagine a process by which an ancient polysynthetic language, with not fully differentiated syntax and semantics, further evolves into a more segmented one, in which typical (for most of us) nouns develop from the anterior protagonist list of the sentence, typical verb develops from the posterior "action" part, and typical adverbs, adjectives, and pronouns scatter from what is in between, releasing also some nouns frozen into the "verb." By backtracking we can descend to more primitive forms in which noun and verb, the primary and the applied components, are not differentiated. In this way we can see relation between Navajo, Swahili, Japanese, and Indo-European languages. All languages are equal under Thought, but some are more equal than others.

The noun classifiers in Chinese and classificatory verb stems in Navajo look like the same linguistic phenomenon connecting the two languages across the Pacific, which is not quite surprising after all if we accept the Dene-Caucasian hypothesis [8]. There is, however, much more similarity between Chinese and Navajo in the very nature of close relation between thought and speech, topicalization, and tightly fused sentence, while classificators seem to be a more ancient phenomenon. Both languages use ideograms: one for writing, the other for speech. Naturally, Chinese needs phonetic elements in its ideograms, which Navajo has *gratis*. Both are poorly equipped for representing new sounds and concepts. But the sparse, lean English-like Chinese needs more of context.

The verb in the Euro-centric sense is always at the very end of the Navajo "verb," where it signifies a very abstract idea of a pattern of change (for example, simple horizontal movement of Direct Object from Subject to Indirect Object, which is called **migration** in chemistry) and a more concrete but abstract enough situation to which it is applied (something in an open container), plus the very elaborate dynamics of the process (for example, tense, mode, and aspect). This is what **nikaah** (to give something in an open container, momentaneous imperfect form, 3rd person singular, if I am not mistaken, in Example means, although this entire Note is a gross simplification. The stem of verb

nikaah is **KĄ́** [7A, p. 294; 7C] , which means an object or substance in an open container. Nevertheless, if one can handle chemical complexity, so one can manage the complexity of Navajo—the language which is an adventure in itself. Chemistry becomes simple in simple cases and so does Navajo. The problems, however, arise when we apply a complex language to a complex reality, in both chemistry and linguistics. The slow communication makes little sense, unless in correspondence between philosophers.

The wonderful worlds of Navajo and Chinese leave open unique windows into human thought, but they should be better left to another opportunity of discussing Nean and kinetic aspects of language evolution.

7. CONCLUDING REMARKS

Although I ignored here the Hymalayas of related literature, one major linguistic movement that starts with a natural-scientific analysis of an event, **before it becomes a thought**, must be mentioned. It is associated with the names of George Lakoff, Mark Johnson, and other enthusiasts of cognitive linguistics, to whom a chemist can feel an instinctive attraction [2B]. As Laura Janda writes,

> From the very beginning, cognitive linguistics has been a refuge for linguists who are intimately acquainted with real language data and have a profound respect for empirical methods [9].

There are, however, even closer kindred souls for a chemist who asks "what if the words were indeed atoms?" Leonard Talmy [10] develops in his cognitive semantics the entire directly observable **physics** of the world from the point of view of its linearization into speech. This seems to me a remarkably innovative and tall idea, to which the chemist alone would not arrive, but could *post factum* formulate it as: **the things have geometrical, physical, chemical and other properties, but they also have a** property

of being told about **in a language.** The languages of Navajo type open a window into the fascinating process of language evolution driven by the general **kinetic properties** of the world, noticed first by René Thom [11, 2B].

My first main idea is that **all** components of human thought, i.e., primary and applied ideas, are connected, although some are more connected than others. The connections may vary over the short time while the thought resides in the mind before either being abandoned, or dumped to memory, or expressed in speech or action. This assumption of an attention window strongly limits the size of the thought. *Ars longa, vita brevis*. The language of art, science, and culture may be long, but we do not have time or memory space for long thoughts.

How much chemistry is in this strange conjecture? Quite a lot, I believe. All atoms interact at a short distance and either attract or repel each other, with repulsion predominant at a very close distance. This is how the proteins take their shape by minimizing the overall strain. I assume that all ideas in the thought are in the neighborhood of each other. **The same is observed in speech** and this is how the sentence takes its shape: "**Speech in observed the is same**" is utterly strained, while "**In speech the same is observed**" is much less so.

The local principles alone are clearly insufficient for ordering modern sophisticated language, which is a result of a long cultural evolution influenced by writing and public oration. It is the language of style and culture, not just of the thought. But they might be sufficient to fill the gap between the protolanguage and the cultivated language of writers and philosophers with its long-distance interaction and central planning of the composition. They might also be sufficient for the spontaneous language stimulated by a fast developing event. Linguistics as a natural science must be the science of time, as chemistry dominated by kinetics is.

There is no contradiction between thinking in language or in a mysterious **language of thought**. When we think in language, for example, internally formulate ("think over")

what we want to say before actually saying that, we simply broadcast the silently prepared and memorized text. When we see a person who is in a danger of being hit by a car and want to warn him or her, we do not have time to think and are in the same position an ancient man was while watching a lion behind the back of his companion hunter.

> Linguistics as natural science cannot have sharp borders. But let us deal only with the problems related to thinking and generating speech. If we decided to deal also with actions, we would come to a firm conclusion that mice think, as all animals do, and this would make some of us feel very uncomfortable while eating them.

My second main idea (partly resonating with the ideas of Sergei Starostin [8]) is that in order to understand language as a natural phenomenon we have (1) to start with very **small** and primitive systems, (2) to see how a **small** evolutionary step can be made, and (3) to apply the procedure repeatedly not only up to present, but also up to a not too remote **future** of language, so that we could check the theory. If this reminds of the method of mathematical induction or of classical Darwinism, I absolutely do not mind. The language acquisition by children does not recapitulate language origin. What both represent is evolution of complexity from simplicity by simplicity.

> WISPERING TO A HORSE: My purpose was to test if I would be able to cross a river over a thin ice from the thought side to the speech side, or, one could say, to circumnavigate the brain from the right to the left hemisphere, in my chemical rubber boots. I am far from my goal and not even sure that I am crossing the right river and even know the right from left. Later I will, probably, give it another try by applying the principles stated here to the lingo-chemical system of *The Tale of Three Little Pigs* , which would supply the component missing here: pragmatics, i.e. the naïve physics of the world. But do I really know what I mean by what I say? As I stated here and in previous e-papers, to understand means to tell somebody else. In a short story by Chekhov, the man tells his troubles to a horse.

REFERENCES

More references can be found in [2].
See also http://spirospero.net/complexity.htm

1. Chalmers, David. *Online papers on consciousness, part 2: Other Philosophy of Mind.* Compiled by David Chalmers. http://consc.net/online2.html

2. Tarnopolsky, Yuri.

   A. *Molecules and Thoughts*. *Molecules and Thoughts: Pattern Complexity and Evolution in Chemical Systems and the Mind* , 2003.
   http://www.dam.brown.edu/ptg/REPORTS/MINDSCALE.pdf
   or: http://spirospero.net/MINDSCALE.pdf

   B. *TIKKI TIKKI TEMBO: The Chemistry of Protolanguage*
   http://spirospero.net/Nean.pdf

   C. *Pattern Theory and "Poverty of Stimulus" argument in linguistics*.
   http://spirospero.net/Poverty of stimulus.pdf

   D. *The Three Little Pigs : Chemistry of language acquisition*.
   http://spirospero.net/3LP.pdf

   E. *Salt: The Incremental Chemistry of Language Acquisition*
   http://spirospero.net/Salt.pdf

   F. Salt 2: *Incremental Extraction of Grammar by Simplistic Rules*
   http://spirospero.net/Salt2.pdf

   G. *Molecular computation: a chemist's view*.
   http://spirospero.net/PTutor.pdf

3. Organic nomenclature
http://www.absoluteastronomy.com/encyclopedia/o/or/organic_nomenclature.htm
   See also any textbook of organic chemistry.

4. Miller, Jonathan. *General Experimental Details*
http://www.jonathanpmiller.com/experimental/
   In: Jonathan Miller's Home Page.
   http://www.jonathanpmiller.com/

5. Grenander, Ulf.

A. 1995. *Elements of Pattern Theory*. Baltimore: Johns Hopkins University Press.

B. 1993. *General Pattern Theory. A Mathematical Study of Regular Structures*, Oxford, New York: Oxford University Press. (Advanced).

C. *Patterns of Thought*. www.dam.brown.edu/ptg/REPORTS/mind.pdf

Watch for updates; see also: www.dam.brown.edu/ptg/**publications**.shtml

6. Halloran, John. The Proto-Sumerian Language Invention Process http://www.sumerian.org/prot-sum.htm]

7. Navajo Language

A. Young, Robert W. and Morgan, William, Sr. 1992. *Analytical Lexicon of Navajo*. Albuquerque: University of New Mexico Press.

B. McDonough, Joyce M. 2000. How to use Young and Morgan's "The Navajo Language." In K. M. Crosswhite & J. S. Magnuson (Eds.), U*niversity of Rochester Working Papers in the Language Sciences, 1 (2),* 195-214. www.bcs.rochester.edu/cls/f2000n2/mcdonough.pdf

C. *Analytical Lexicon of Navajo*. Online, under construction . Stem KÂ. http://www.speech.cs.cmu.edu/egads2/navajo/entry?K%c2

8. The Dene-Caucasian Hypothesis (John Bengtson, Sergei Starostin). Santa Fe Institute. http://ehl.santafe.edu/denecauc.htm . See also the entire EHL project.

9. Janda, Laura. Cognitive Linguistics. www.indiana.edu/~slavconf/SLING2K/pospapers/janda.pdf

10. Talmy, Leonard. 2000. *Toward a Cognitive Semantics*. Cambridge, Mass. The MIT Press,. http://linguistics.buffalo.edu/people/faculty/talmy/talmyweb/TCS.html

11. Thom, René. 1975. *Structural Stability and Morphogenesis: An Outline of a General Theory of Models.* Reading, Mass.: W. A. Benjamin, Inc.

yuri@ids.net

First published: May 28, 2005

Last revision: July 19, 2005

What if the words were indeed atoms?