

# LexiClone

*lexical cloning system*



## Differential Linguistics at NIST TREC.

Ilya Geller

[igeller@lexiclone.com](mailto:igeller@lexiclone.com)

### Abstract.

In the course of carrying out NIST TRECs I created and tested a computer program for textual information searches, based on ‘understanding’ the meanings of words in texts. The computer using the program ‘understands’ not only the abstract, standardized meanings of the words in the text, but the specific, concrete meanings given to those words by the author(s) of the texts. In this article I attempt to bring the language I used to create the algorithm of the program in line with the generally accepted, formalized language of mathematics. (Doing this I must apply the philosophy and metaphysics of Cynicism.)

### Axiom 1. Words exist.

**Definition 1.** I understand a word in any given language to be a combination of letters in that format in which the word appears in print in a generally accepted dictionary of that language. That combination of letters by which the word is fixated in the dictionary is recognized as the ‘normal form’ of that word, to which all ‘non-normal’ forms of the given word can be reduced; by a ‘non-normal’ form of a word I mean a form which arises from adding prefixes, suffixes, endings, etc., to the normal form of the word; or a form resulting from the introduction of a grammatical error into the word.

Use of the dictionary of a language allows one to present each word in numerical form. Differential Linguistics thus works with numbers; and the system for reducing non-normal forms of words to their normal forms can be seen as a system for reducing words to numbers.

**Definition 2.** The meaning of a word is how the word *is used* and what the word *is*.

**Definition 3.** Any word taken separately in its normal form is a ‘non-predicative definition’. I have called combinations of normal forms of words – nouns/pronouns-verbs-adjectives – ‘predicative definitions’.

**Note. Noam Chomsky.** In 1957 Noam Chomsky [1] proposed calling the combinations of words that convey the meaning of a sentence ‘kernel sentences’. But I have preferred to follow an immeasurably more ancient tradition which had its beginning with Aristotle, and to call such combinations ‘predicative definitions’ (if they are reduced to their normal forms.)

**Definition 4.** I understand only the normal form of a word to be a non-predicative definition; where a non-predicative definition and an abstraction/universal [6; ‘The World of Universals’] are the same thing.

I claim that any non-predicative definition has all words’ meanings.

**Clarification. Philosophy.** I have chosen, as an intellectual basis for my program, the philosophy of Cynicism, which I see as superseding the philosophy of Idealism.

---

**Acknowledgement.** I thank Dr.Inna Rozentsvit, NYU and Member of the Russian Academy of Medicinal Sciences Dr.Gennady Sukhikh for saving my life.

Also, I am immensely grateful to my friend Aleksandr Syrkin, who always helps me. Without him this article could not be written.

I suppose that Cynicism, as a collection of dogmas in written form, was originally created by the Biblical authors Ecclesiastes and Jeremiah who, I think, were opposing Plato. I consider Cynicism to be based on the notion that the Ideal is single, doesn't change and therefore doesn't exist in time – the (only) Ideal exists only in the immutability of eternity, in timelessness. I am certain that Idealists Plato and (in some sense) Aristotle – as well as their followers Hegel, Bradley and Russell – supposed that there exists a multiplicity of ideals, that ideals exist in time and not in time and can be distinguished one from another [5,6,11]. (Idealists have never explained how one could make distinctions between things that were the same – absolutely identical. I mean this: how one could make distinctions between two abstracts? [20,21] Hegel pretended that the question didn't exist – see his 'The Science of Logic': he supposed that *nothing* exists in time as well as *beings*.)

Having applied the concepts of Cynicism to Linguistics I have come to the aforementioned conclusion, stated in Definition № 13, that the normal form of a word is the abstraction – that is, it has an indefinitely large and in no way distinguishable number of meanings as long as the word hasn't been combined with other words and/or given a non-normal form.

For instance, a word '*ggffrrtte*'. In its normal form it means everything and nothing at the same time – unless it is explained. Hegel said about this phenomenon: '... pure being is the *pure abstraction* ... which when taken immediately, is equal *nothing*. From this... a definition of the Absolute followed, that it is *nothing*... Hence, the truth of being and nothing alike is the *unity* of both of them; this unity is *becoming*' [5, p.140-141]' To possess a meaning of the word '*ggffrrtte*' the word should become! I think that the inclusion of the normal form of a word into a structure in combination with other words and/or its modification into a non-normal form transforms the [abstract] word into a concrete word with a concrete meaning through its becoming<sup>1</sup>.

**Note. Russell's Non-Predicative Definition.** Bertrand Russell introduced the notion of a 'non-predicative' definition, in which what is to be defined is brought in through its relation to a class of which it is an element [6,7,8,9]. For example: 'the set of all sets that are not elements of themselves'.

Formally, linguistically, the given affirmation – 'the set of all sets that are not elements of themselves' – is a combination of sets of words in normal and non-normal form, intended to clarify the meaning of the word 'set'. The word 'set' by itself, however, in its normal form and not in combination with other words, can implicitly carry the meaning it has in 'the set of all sets that are not elements of themselves' together with many other meanings<sup>2</sup>. And the same with the word '*ggffrrtte*' till it's explained!

<sup>1</sup> 'Indeed no sooner have we distinguish the two words than it becomes necessary to consider their relations.' [2, p.100].

<sup>2</sup> For example, the **Merriam-Webster Online Dictionary** gives the following list of meanings for the word 'set', taken only as a noun:

Main Entry: set

Function: noun

1 a : the act or action of setting b : the condition of being set

2 : a number of things of the same kind that belong or are used together <an electric train set>

3 a : mental inclination, tendency, or habit : BENT <a set toward mathematics> b : a state of psychological preparedness usually of limited duration for action in response to an anticipated stimulus or situation <the influence of mental set on the effect experienced with marijuana>

4 : direction of flow <the set of the wind>

5 : form or carriage of the body or of its parts <her face took on a cynical set -- Raymond Kennedy>

6 : the manner of fitting or of being placed or suspended <in order to give the skirt a pretty set -- Mary J. Howell>

7 : amount of deflection from a straight line <set of a saw's teeth>

8 : permanent change of form (as of metal) due to repeated or excessive stress

9 : the act or result of arranging hair by curling or waving

10 also sett /set/ a : a young plant or rooted cutting ready for transplanting b : a small bulb, corm, or tuber or a piece of tuber used for propagation <onion sets>

11 or sett : the burrow of a badger

12 : the width of the body of a piece of type

13 : an artificial setting for a scene of a theatrical or film production

14 also sett : a rectangular paving stone of sandstone or granite

15 : a division of a tennis match won by the side that wins at least six games beating the opponent by two games or by winning a tiebreaker

16 : a collection of books or periodicals forming a unit

17 : a clutch of eggs

18 : the basic formation in a country-dance or square dance

19 : a session of music (as jazz or dance music) usually followed by an intermission; also : the music played at one session

20 : a group of persons associated by common interests

21 : a collection of elements and especially mathematical ones (as numbers or points) -- called also class

22 : an apparatus of electronic components assembled so as to function as a unit <a television set>

To make an analogy with Set Theory: if it is given that the normal form of a word has a countable number of meanings  $N$ , then when that word is included in a combination of many words that set of meanings is reduced to the dimensions of its intersection with the sets of meanings of other words – to the set  $M$ ,  $M \subset N$  ( $M$  is a subset of  $N$ , the power (number of elements) of  $M$  is less than or equal to the power of  $N$ ).

**Axiom 2.** There exists a countable and limited number of parts of speech.

**Definition 5.** Each part of speech explains not what the word *is*, but only how the word *is used*.

Any normal form of a word can belong to one or to several parts of speech.

**Postulate 1.** In most cases, it is only possible to identify the part of speech to which a given word belongs by analyzing the combination of the normal or non-normal form of the given word with the normal and/or non-normal forms of other words - with the given word present explicitly or implicitly in such combinations<sup>3</sup>. In a few cases one can identify the part of speech to which a given word belongs even without analyzing the combination of the normal and/or non-normal form of the given word with other words<sup>4</sup>.

**Definition 6.** Words and their combinations always appear as parts of a sentence<sup>5</sup>, and sentences *always* form a paragraph<sup>6</sup>.

**Axiom 3.** There is Reality<sup>7</sup>. Words, sentences and paragraphs are used for the describing of Reality.

Reality is everything that is and is not. Also, I am not interested in whether the unicorn really lives or in how long its horn is: it is a part of Reality, since Reality is everything that is and is not [6,7,8,9,16.17.18].

**Postulate 2.** Reality is always and continuously changing.

I am unaware of any unchanging Reality.

**Definition 7. Context.** I understand context to be the description of concrete, named parts of Reality and of what happens to them.

Context can only be provided by a sum of combinations – always no less than one combination – of normal forms of words, extracted from a sentence of a paragraph of a text; where the normal form of words is arrived at by the reduction of non-normal forms to normal. Moreover, such combinations must always conform to the structure of the following triad: substantive (noun), verb, adjective.

**Definition 8. Subtext.** I understand subtext to be the description of unnamed and unnamable parts of Reality and of what happens to Reality and its parts. Subtext can only be provided by a sum of combinations – always no less than one combination - of normal forms of words – combinations of no less than three parts of speech – pronoun, verb and adjective.

A method and system for extracting context and subtext from texts/paragraphs is described in the article ‘The Role and Meaning of Predicative and Non-Predicative Definitions in the Search for Information’ [2]. The system consists of

- a) The reduction of the non-normal forms of the words used in texts and paragraphs to their normal forms

---

23 : a usually offensive formation in football or basketball

24 : a group of a specific number of repetitions of a particular exercise

<sup>3</sup> By the implicit presence of a word I mean that the word is not stated directly in a given combination but is implied to be.

<sup>4</sup> Such a word is usually a proper Name or an appellation - a noun, although it could be any other part of speech. Such a word has a set of meanings  $N$  which has no subsets; the power of  $N$  is equal to 1.

<sup>5</sup> The **Merriam-Webster Online Dictionary** defines a sentence as follows:

Main Entry: sen-tence

Function: noun

4 a : a word, clause, or phrase or a group of clauses or phrases forming a syntactic unit which expresses an assertion, a question, a command, a wish, an exclamation, or the performance of an action, that in writing usually begins with a capital letter and concludes with appropriate end punctuation, and that in speaking is distinguished by characteristic patterns of stress, pitch, and pauses b : a mathematical or logical statement (as an equation or a proposition) in words or symbols

<sup>6</sup> The **Merriam-Webster Online Dictionary** presents the meaning of the word ‘paragraph’ as follows:

Main Entry: par-a-graph

Function: *noun*

1 a : a subdivision of a written composition that consists of one or more sentences, deals with one point or gives the words of one speaker, and begins on a new usually indented line b : a short composition or note that is complete in one paragraph

<sup>7</sup> The **Merriam-Webster Online Dictionary**:

Main Entry: re-al-i-ty

Function: *noun*

1 : the quality or state of being real

2 a (1) : a real event, entity, or state of affairs <his dream became a *reality*> (2) : the totality of real things and events <trying to escape from *reality*> b : something that is neither derivative nor dependent but exists necessarily

- in reality : in actual fact

b) And the compilation of the combinations of those words within the sentences of texts and paragraphs.

The method thus consists of the summarization of the combinations of normal forms of words found in texts and paragraphs (taken as sums of sentences) in order to establish the contexts and subtexts of texts and their paragraphs (see US Patent № 6,199,067).

I am certain that the context and subtext of a paragraph provide a very limited set  $R$  of meanings for every word used in the paragraph, where  $R \subset M$  [ $R$  is a subset of  $M$  which has a power less than or equal to that of  $M$ , where  $R \subset M \subset N$ ].

**Definition 9.** A paragraph is devoted to the description of a limited number of parts of Reality: a paragraph, as distinct from the sentences that compose it, must in the overwhelming majority of cases have an absolutely simultaneously defined context and subtext, which allow one in most cases to understand without ambiguity the meaning of the words used in the paragraph.

If a paragraph does not have a simultaneously defined context and subtext, it means that an error has crept into the paragraph and/or it should be linked to (an)other paragraph(s) of the text.

**Observation 1.** Context and subtext can only be provided by a sentence, a paragraph, or a text.

**Note to Observation 1.** In order to provide context and subtext one needs a minimum of one combination of normal forms of words. Such a combination of normal and/or non-normal forms of words is a sentence. And a paragraph and a text are a sum of some/many sentences.

**Observation 2.** It is not always possible to determine to which part of speech a word belongs by examining only one single sentence in which it is explicitly or implicitly present.

**Note to Observation 2.** A sentence may consist of only one word which is neither a proper Name nor an appellation - i.e., it is what I call a 'single-part sentence'. In that case it is impossible to examine the combination of words that would allow one to identify the part of speech to which the word belongs, since there is no combination of words in the sentence.

For example, if someone creates the sentence, 'Red', one cannot tell to which part of speech the word used in the sentence belongs. 'Red' could refer to a colour, or be a pejorative term for a Communist, or it could be a proper Name or an alias. In the first case the word 'red' would be an adjective, in the others it would be a noun.

**Observation 3.** The definition of how a word *is used* – to which part of speech it belongs – is most often made possible by an examination of the paragraph in which the word is used.

**Note to Observation 3.** Indeed, since a paragraph is a set of sentences – no less than one sentence – then a paragraph makes it possible to define the part of speech to which a word belongs.

**The Function of the Description of Reality.** Description in words is dependent on changing Reality.

I have come to the conclusion that as Reality changes the description of it always and inescapably changes as well, reflecting the changing of Reality itself. For example, if someone dropped a cup of tea on the floor and it broke, spilling the tea, then a description of the broken cup and the spilled tea would be entirely different from a description of the cup before it fell on the floor.

Moreover, out of the given set of all possible subjective descriptions of Reality  $E$ , one can relate to described Reality – let's call it  $x$  – only one subjective description of Reality, designated as  $y=f(x)$ . One can then say that for the set of descriptions of Reality  $E$  a function of description of Reality is provided in the terms

$$y = f(x), x \in E$$

where  $E$  is the field of subjective definition, the set of all possible states of Reality.

**Postulate 3.** Only subjective descriptions of Reality exist.

**Note. Subjectivism**<sup>8</sup>. The function of description of Reality in words is always subjective, since only a subject (the observer) is in a position to describe Reality in words. In any case, I am unaware of any descriptions of Reality created directly or indirectly by an object, and not by a subject. I am certain that even if objects exist that are capable of describing Reality in words, they were created and/or taught to describe Reality by a human being.

The function of the description of Reality can be provided analytically, if one considers proven the hypothesis of the presence in every subject of an individual and limited set of lexical habits [3], the existence of which is determined simultaneously by the

---

<sup>8</sup> The **Merriam-Webster Online Dictionary**:

Main Entry: sub-jec-tiv-ism

Function: *noun*

**1 a** : a theory that limits knowledge to subjective experience **b** : a theory that stresses the subjective elements in experience.

Aesthetic and Ethical components [4] of the subject's mind. Indeed, if one knows which words and combinations thereof must inescapably be used by the subject in describing Reality (and its parts), one can analytically provide the function of description of Reality by the given subject. For this one must know

1. The subject's emotional relation to Reality and its parts; knowledge of which is transmitted by the Ethical component of the subject's mind,
2. And it is also necessary to know about the subject's own knowledge of Reality and its parts: this knowledge is the Aesthetic component of the subject's mind.

**Clarification.** Words have only subjective meanings.

**Note. Grammar.** In order to extract/reproduce the meaning of the subjective function of description of Reality at a given moment of time in the form of combinations of normal and non-normal forms of words Grammatically coordinated in sentences and paragraphs (that is, in the form of a text) it is necessary to have an understanding of the laws of Grammar. At the present moment I don't have that necessary understanding of Grammar.

**Note. The Ancients.** I take the limit of the function of description of Reality to be silence, following in this the teaching of the ancient the pre-Platonic Pythagoreans and Hebrews. (I consider Plato and Aristotle, as well as Ecclesiastes and Jeremiah, to have been the heirs of the Pythagoreans. Or, Pythagoras and the pre-Platonic Pythagoreans were the heirs of Ecclesiastes (and, perhaps, Jeremiah), I don't know.)

**Definition 10.** The limit of the function of description of Reality is silence - the absence of the description of Reality. In other words, I consider the function of description of Reality to be a differentiable function for which a limit can be defined. Indeed, there exists a derivative function of description of Reality  $y'=f'(x)$ : it is the limit of the relation of changes in the description of Reality to Reality itself. If the changes in Reality are vanishingly small, then the changes in the description of Reality are vanishingly small. If Reality does not change, then the description of Reality does not change.

$$y' = f'(x) = \lim_{\Delta x \rightarrow 0} dx/dy = \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}$$

where  $dx$  designates changes in Reality,

where  $dy$  designates changes in the function of description of Reality.

In other words, if Reality is unchanging, there is no function of description of Reality: it has attained its limit in silence.

**Observation 4.** The function of description of Reality is continuous.

**Note to Observation 4.** In response to the slightest change in Reality the function of description of Reality changes slightly but inescapably. Since Reality – the argument of the function – is always and continuously changing, then the function of its description is always and continuously changing. The same: it is enough to scoop up some water from the Atlantic Ocean in a glass for the description of the ocean to change at once, however insignificantly.

I base the following speculations on my unshakable certainty that a human being can operate only in terms of the derivatives from the function of description of Reality, divided by parameters<sup>9</sup>. By analogy with mathematics, where the derivative of a function is established on a point, I establish the derivative of the function of description of Reality in an instant of time.

**Postulate 4.** The second derivatives from the function of description of Reality are discrete.

**Definition 11. A derivative from the function of description of Reality.** I understand a derivative from the function of description of Reality to be descriptions of Reality as Reality changes striving for a minimum, in an infinitely small interval of time.

**Definition 12. The first derivative from the function of description of Reality.** The first derivative from the function of description of Reality is the normal form of (a) noun(s) describing changes in Reality in an infinitely small interval of time.

**Definition 13. The second derivative from the function of description of Reality.** The second derivative from the function of description of Reality is a paragraph.

**Note. Time.** Time, as I conceive it, is a parameter used by the subject to divide continuity into fragments for description – into a countable (discrete) number of descriptions. In the article 'LexiClone Inc. and NIST TREC' I supposed that Reality exists in eternity and Truth exists in infinity – where eternity is closed/completed time, and infinity is continuing time [2]. This means that taking an

---

<sup>9</sup> A parameter is a certain quantity, values of which are used to distinguish elements of a given set from one another.

interval of time as a parameter allows one to describe Reality in terms of a finite set of descriptions. But Truth, by reason of its incompleteness, cannot be described in terms of a finite set of (discrete) descriptions. (Future cannot be predicted.)

But it is precisely time which creates the parametric link between the function of describing Reality and Reality itself, time which parametrically sets the function of describing Reality in relation to the function of change in Reality itself.

**Definition 14. The prototype at the integration of a paragraph into the first derivative from Reality.** A group of verbs and adjectives is that part of a prototype which appears and is displayed when a paragraph is integrated into the first derivative from Reality<sup>10</sup>.

**The meaning of a word.** The prototype of a paragraph gives one a sufficient and necessary understanding of a word's meaning – of how the word *is used* and what the word *is*. Given below is a small fragment of the prototype for the preceding paragraph which illuminates the meaning of all the words in the paragraph.

reality - paragraph - derivative

reality - part - derivative

reality - form - derivative

reality - result - derivative

reality - be - derivative

description - paragraph - derivative

description - part - derivative

description - form - derivative

description - result - derivative

description - be - derivative

function - paragraph - derivative

function - part - derivative

function - form - derivative

function - result - derivative

function - be - derivative

reality - paragraph - first

reality - paragraph - into

reality - paragraph - integrated

reality - part - first

reality - part - into

reality - part - integrated

reality - form - first

reality - form - into

reality - form - integrated

reality - result - first

reality - result - into

reality - result - integrated

reality - be - first

reality - be - into

reality - be - integrated

I am certain that the meaning of a word can be understood only if the paragraph provides frequent repetition of a predicative definition containing that word. Sometimes a single paragraph may not contain any repetition of the predicative definition which

---

<sup>10</sup> Substantives (nouns) provide the proper Names by which the subjects-and/or-objects of the paragraphs of a text, as parts of Reality, are known, while the combinations of verbs and adjectives describe what happens with those parts of Reality. '...adjectives and common nouns express qualities or properties of single things, whereas prepositions and verbs tend to express relations between two or more things.' [2, p.94].

conveys the relevant meaning of a word. In that case it is necessary to read several paragraphs – before and/or after the given one – to provide enough repetitions of the same predicative definition, which will allow one to understand the relevant meaning of the word.

It means that if a person (and my program as well) begins reading a text from its end, it could be necessary to somehow establish the context and subtext of what was in the book earlier, in order to understand the meanings of the words. One (and my program as well) can only gain such knowledge by reading further and/or going back a few pages/paragraphs.

**Definition 15. Summary.** The prototype of the second derivative from the function of describing I call ‘summary’. A summary is a sum of lexical habits.

**NIST TREC QA.** For NIST TREC QA I worked out a methodology of text paragraph fusion and subsequent restoration of the sentences of the paragraphs.

Paragraph fusion consists in uniting several paragraphs of a text that are linked contextually and/or subtextually and/or mechanically. Restoration of the sentences of the paragraphs – by context – is generally achieved by adding the predicative definition ‘it-be-in’ to all the sentences needing restoration.

**Paragraph Fusion.** If a single paragraph of a text does not provide a trustworthy – that is, identical – understanding of the ‘generally accepted’ slang used at its creation<sup>11</sup> – usually because the paragraph lacks even one complete predicative definition<sup>12</sup> – then in order to understand the meanings of the words of that paragraph it is necessary to be acquainted with the preceding and/or following and/or other paragraph(s) of the text. In other words, an ‘untrustworthy’ paragraph of this kind is linked contextually and/or subtextually and/or mechanically to (an)other paragraph(s) of the text. (For example, if a paragraph consists of only one monosyllabic sentence, containing only one word, then that paragraph can necessarily be fused with the preceding paragraph. The same procedure may be necessary if the paragraph is part of a dialogue and begins with a hyphen [quotation mark]. It is also necessary to fuse sentences that represent a question and answer in the text – a kind of crossword. And so on in the same vein.)

By fusing paragraphs of text my program creates paragraphs containing summaries of predicative definitions necessary to understanding the meanings of the words in them – paragraphs created within the framework of a certain ‘generally accepted’ slang.

**Restoration of Sentences of Paragraphs.** Monosyllabic sentences/sentences lacking (even one) full predicative definition can and must be restored (after necessary, preliminary paragraph fusion).

Monosyllabic sentences are (usually, almost always) sentences that deal with Reality. In other words, they contain either a question or an affirmation about something finished – that is, about what was, is, or will be. (In my article ‘LexiClone Inc. and NIST TREC’ [2] I affirm that Reality exists in completed times). I thus have all the necessary grounds to add to monosyllabic sentences the predicative definition ‘it-be-in’, as:

- a) containing in itself an affirmation that the abstract ‘it’ was, is, or will be:
- b) having in its structure the most frequently recurring words in the summary of any text, almost any paragraph and almost all sentences: the verb ‘be’ and the adjective [preposition] ‘in’. (NIST TREC is interested exclusively in the retrieval of contextual information, since before the publication of my article [3] there was no awareness of even the existence of the subtext. In the case where a search by subtext would be needed I would add the predicative definition ‘i-be-in’.)

By restoring sentences my program recreates sentences that contain tautological<sup>13</sup> answers to queries put in the course of NIST TREC QA. In addition to sentence restoration and paragraph fusion a number of other steps must also be taken to find answers to NIST TREC QA queries.

---

<sup>11</sup> Where a slang is understood to be a collection of predicative definitions with a unique (within the limits of statistical error) relative frequency of use; where such frequency is characteristic of only one, strictly defined area of human activity. Statistical error in the frequency of use of a slang's predicative definitions is the result of subjectivity in their use. For a more detailed introduction to the definition of ‘slang’ cf. my article ‘The Role and Meaning of Predicative and Non-Predicative Definitions in the Search for Information’, [3].

<sup>12</sup> A predicative definition is judged to be complete if it contains in its structure at least three words belonging to parts of speech necessary for the creation of either a contextual or a subtextual predicative definition. For a more detailed introduction to the rules for constructing predicative definitions cf. my article ‘The Role and Meaning of Predicative and Non-Predicative Definitions in the Search for Information’, [3].

<sup>13</sup> **Merriam-Webster Online Dictionary** defines tautology as ‘needless repetition of an idea, statement, or word’. See also my article *LexiClone Inc. and NIST TREC* [2] and Ludwig Wittgenstein, *Tractatus Logico-Philosophicus*, Hint #4.4.

**Note. Lexical Clone.** As I stated earlier in my article [2] my program seeks an answer to a query in the AQUANT collection of texts after having already found one in a smaller and specific database in a text created in the ‘generally accepted’ slang of the query. Moreover:

- 1) Having found an answer to a NIST TREC QA query in the smaller and specific database,
- 2) The program extracts from the summary of the predicative definitions of such a text a small number of predicative definitions that are tautological to the given query, and that in fact already contain the answer to the query in the format that NIST TREC QA requires (a single word or very few words). Then my program seeks in the AQUANT collection a sentence with, again, predicative definitions that are tautological to the ones already found in the smaller and specific database.

A text containing in itself the answer to a NIST TREC QA query found in the smaller and specific database could be called the prototype of a Lexical Clone, the imprint of a person's consciousness. That is, if such a text could (hypothetically) be read by the person seeking the answer to that query, it would become a part of his consciousness (and/or would in some way influence his consciousness). It would influence it in the sense that the slang of such a text would become known and intelligible to that person. And in a sum with a set of texts (created within the framework of a multiplicity of slangs), a Lexical Clone, as a summary of predicative definitions from a set of texts and/or their paragraphs, can be seen to be a kind of print of a person's consciousness.

**The Goal of NIST TREC QA.** The goal of NIST TREC QA is to locate, in an enormous quantity of texts, cursory and insignificant references to the object of a search. In other words: NIST TREC QA presupposes a primordial, unambiguous, extremely focused understanding of the words in its query, within the framework of a certain ‘generally accepted’ slang which must be ‘known’ to the computer. Such an effect is attained through the formulation of the query, in which one uses the predicative definitions most frequently employed in the aforementioned ‘generally accepted’ slang.

**Clarification. Cynical Metaphysics**<sup>14</sup>. Cynical Metaphysics is intended to establish a metaphysical (mathematical) foundation for my Differential Linguistics.

I think that it is clear that an Idealist approach leads to abstracting from the inner substance of the objects-and-subjects of the surrounding world, as well as to abstracting from the differences in their form (in the spirit of the *polytheism* of Plato): it leads to the standardization of the objects-and-subjects into ideals, into what are traditionally called ‘material points’. (Didn't Plato indeed have to find a way of unifying the multiplicity of coexisting gods and objects-and-subjects? Among the best-known examples of such Platonic unification is the Mechanics of Isaac Newton).

When applied to the theory of Information Retrieval, the aforementioned Idealist principles indicate that different ‘material points’ need the same information (no one knows for what).

Cynicism, however, holds to another point of view: the abstract ideal is single and exists changelessly outside time<sup>15</sup> (in the spirit of Biblical, The Bhagavad-Gita's and Upanishads' *monotheism*). In other words, the ideal cannot be experienced sensually and, therefore, does not exist/ exists not in time. The objects-and-subjects of the surrounding world, however, exist in time and can be experienced sensually, only and only through interactions<sup>16</sup>. The objects-and-subjects of the surrounding world are a form of topological ‘points of accumulation’ [22, 23]: there is no way of definitely and absolutely separating the objects-and-subjects of the surrounding world, as closed sets consisting of their parts<sup>17</sup>. For me, as a Cynic, any object-subject presented to us by our perceptions of the Universe is an open, continuously changing set of the parts comprising it; such a set interacts continuously with all the objects

---

<sup>14</sup> **Merriam-Webster Online Dictionary** thinks that metaphysics is the system of principles underlying a particular study or subject.

<sup>15</sup> Bertrand Russell held to the same point of view: ‘But universals ... *subsist* or *have being*, where ‘being’ is opposed to ‘existence’ as being timeless. The world of being is unchangeable, rigid, exact, delightful to the mathematician... and all who love perfection more than life.’ [6, p.100] In his book Bertrand Russell makes a more detailed analysis of the problem of the nature of the ideals/ universals in the chapter ‘The World of Universals’ (to which I address the interested reader).

<sup>16</sup> An observer always interacts with what is observed, and in general action without interaction is unthinkable – and this is the beginning of my New Mechanics. [12,13,14,15,19]

<sup>17</sup> From the point of view of my interpretation of Cynicism, the boundary problem in relation to the objects-and-subjects of the Universe has no solution and cannot have one. The Cynical path to the standardization of the objects and subjects of the surrounding world is a statistical one: the parts of the Universe can be distinguished only quantitatively, within the limits of statistical error [7,16,17,18,20,21]. ‘Material points’, however, are the closed set(s) [19].

and subjects of the surrounding world, without exception. That is, Cynicism solves the problem of unifying the objects-and-subjects of the surrounding world by approximately (within the limits of statistical error) evaluating the number of the parts of sets<sup>18</sup>.

**Clarification. Cynical Metaphysics: ‘Virtue Of Concurrence’.** Following in the footsteps of Aristotle, I affirm that there is a ‘virtue of concurrence’<sup>19</sup> that makes objects-and-subjects unite in a certain open set/ object-or-subject as long and in so far as such an object-or-subject can make the parts ‘better’ than they were before; where the best is the goal and the result of the development of all that exists<sup>20</sup>. (The comment about ideal and goal being indistinguishable is (in some sense) mine; Aristotle might not think that way). The non-existence of ideals provokes ‘virtue of concurrence’ to be.

When applied to the theory of Information Retrieval, the aforementioned Cynical principles indicate that different points of accumulation – which are often to be called people – need different information about how to become ‘better’. Thus for me the role of Cynicism in Information Retrieval is to bring one to the understanding that different information is needed by different points of accumulation so that they can become ‘better’. (And the mentioned in my article [2] Compatibility helps to find ‘tautological’ information that makes one ‘better’ – subjectively ‘better’, the ‘best’.)

**Note. Linguaprints.** I consider the only way to necessary personalization in Information Retrieval to be the extraction of a subject’s linguistic habits, which provide a rich description of the subject’s inner world, from within. Why? I see no other way of somehow empirically fixating human consciousness outside the subject. As I have demonstrated [2], the collection of a person’s linguistic habits is a unique external representation of an individual’s consciousness, similar to fingerprints (I’ve called such collections of individual linguistic habits ‘linguaprints’).

**Note. The Facts.**

1. During NIST TREC QA 2003-5 my program searched for paragraphs.
2. The fact that the largest US companies involved in information retrieval and advertising on the Internet moved on to
  - a) The personalization of information retrieval and advertising and
  - b) Contextual information retrieval and advertising

seems to confirm that I took the right path many years ago (1993-97) embracing Cynicism precisely when everything in science and Information Retrieval appeared so stable, permanent and clear.

3. This year I used my own PC: Athlon 700MGz, 128RAM and 40Gig of memory; Windows XP. However, it’s not enough.

**Hypothesis I.** A text is a third derivative (?) from the function of description of Reality.

**Hypothesis II.** The lower and upper limits of a paragraph’s integral are determined by a (search?) request for information how to become ‘better’ and, finally, the ‘best’. I think that so called ‘keywords’ (non-predicative definitions, nouns), actually, establish *the region under the curve  $y = f(x)$  between synonyms (?) of the keywords.*

**Conclusion.** I brought the language I used to create the algorithm of my program in line with the formalized language of mathematics.

I explained why I have come to the conclusion that the role and significance of a paragraph are now clear: a paragraph is a second derivative from the function of description of Reality, usually having a context and a subtext. The context and subtext of a paragraph as a rule allow one to understand the meaning of every word included in the paragraph. If they do not – it could be necessary to read several paragraphs – before and/or after the given one. This knowledge of the meanings of words allows my program to ‘understand’ texts.

Also, I made clear why I think that Informational Retrieval cries out for personalization: I think that different people seek different information about how to become ‘better’ and the ‘best’.

**References.**

[1] Chomsky, N. 1957: *Syntactic Structures*. Mouton de Gruyter; Reprint edition (June 1, 1975)

---

<sup>18</sup> Ideals are presented as the reason for the Force of interaction for/ between objects-and-subjects to exist: ideals are considered as the defect of the Universe.

<sup>19</sup> ‘...clearly to say that something comes to be out of what is not, is to say that it does so out of what is not, as something which is not... We too say that nothing comes to be simply out of what is not; but that things do come to be in a way out of what is not, namely by virtue of concurrence’. [11, 191<sup>β</sup>, 5-15]

<sup>20</sup> ‘...for the end should not be just any last thing, but the best.’ [11, 194<sup>α</sup>,30].

- [2] Geller, I. 2004: *LexiClone Inc. and NIST TREC*. The Thirteen Text REtrieval Conference (TREC 2004), National Institute of Standards and Technology (NIST). <http://trec.nist.gov/pubs/trec13/papers/lexiclone.qa.pdf>
- [3] Geller, I. 2003: *The Role and Meaning of Predicative and Non-Predicative Definitions in the Search for Information*. The Twelfth Text REtrieval Conference (TREC 2003), National Institute of Standards and Technology (NIST): 386-391.
- [4] Geller, I. 2004: *Summarization by Means of the Lexical Cloning Method*. The «Web Journal of Formal, Computational & Cognitive Linguistics». <http://fccl.ksu.ru/issue6/pvt2.pdf>
- [5] Hegel, G. 1991: *The Encyclopedia Logic*. Hackett Publishing Company, Inc., Indianapolis.
- [6] Russell, B. 1984: *The Problems of Philosophy*. Hackett Publishing Company, Indianapolis/Cambridge.
- [7] Russell, B. 1919: *Introduction to Mathematical Philosophy*. George Allen and Unwin Ltd, London.
- [8] Russell, B. 1908: 'Mathematical Logic as Based on The Theory of Types' in *Logic and Knowledge*. Essays 1901-1950, Robert Marsh, ed. The Macmillan Company, NY, 1956.
- [9] Russell, B. 1931: *Review of The Foundations of Mathematics and other Logical Essays*. Mind 40. 476-482.
- [10] Voorhees, E. M. 2003: *National Institute of Standards and Technology (NIST). Overview of TREC 2003*. The Twelfth Text Retrieval Conference (TREC 2003).

**Some Extra References.**

- [11] Aristotle. 1995: *A new Aristotle reader*. Princeton University Press, Princeton.
- [12] Bohr N. 1962: *Atomic Physics and Human Perception*. Moscow.
- [13] Bohr N. 1992: *Can Quantum Mechanical Description of Physical Reality Be Considered Complete?* Moscow.
- [14] Bohr N. 1992: *Continuity, Determinism and Reality*. Moscow.
- [15] Gribbin J. 1984: *In Search of Schrodinger's Cat: Quantum Physics And Reality*. Bantam, NY.
- [16] Frege, G., 1980: *Philosophical and Mathematical Correspondence*. G. Gabriel, H. Hermes, F. Kambartel, C. Thiel, and A. Veraart (eds. of the German edition), abridged from the German edition by Brian McGuinness, translated by Hans Kaal, Chicago: University of Chicago Press.
- [17] Frege, G., 1974: *The Foundations of Arithmetic*. J. L. Austin (trans.), Oxford: Basil Blackwell.
- [18] Frege, G., 1967: *The Basic Laws of Arithmetic*. M. Furth (trans.), Berkeley: University of California.

**Some Unconventional References.**

- [19] Geller I. 2005: *Principles of a New Mechanics (re-writing, 2005)*. [http://lexiclone.com/Principles\\_of\\_a\\_New\\_Mechanics.pdf](http://lexiclone.com/Principles_of_a_New_Mechanics.pdf)
- [20] Geller I. 1996: *Clarifications C To Appendix #10a. Concerning Pure Lie About Pure Truth And Pure Truth About Pure Lie: Arithmetic And Dialectic Of Language; Language Of Arithmetic And Dialectic*. <http://lexiclone.com/Ltruth.Pdf>
- [21] Geller I. 1997: *A Sketch Concerning The 'New' Foundation Of Science After Physics*. [http://lexiclone.com/ARTICL\\_BODY.html](http://lexiclone.com/ARTICL_BODY.html)

**Mathematical References.**

- [22] Aleksandrov P., 1977: *An introduction in the theory of sets and general topology*. Moscow.
- [23] Lyapunov A., 1953: *Selected Works. On the Shape of Celestial Bodies*. Moscow-Leningrad.