Towards a Model of Language Understanding

Marius Crisan

“Politehnica” University of Timisoara, Department of Computing and Software Engineering, Bd. V. Parvan 2, RO-300223 Timisoara, Romania
Phone: +40-256-403212, Fax: +40-256-403214. E-mail: crisan@cs.utt.ro

Abstract: The paper is an attempt to outline a hierarchical model of language apprehension based on an extension of Language of Thought Hypothesis (LOTH). Several arguments are presented which show that language being incomplete has limitations in representing both the reality and the mental states. Therefore, postulating LOTH similar with a conventional language is fallacious. Nonetheless, if language is related with thought, language properties would have to have a causal root in the functioning of the mind. This controversial issue is discussed in relation with the possibility of using Zipf’s law for identifying a deeper causal law at the level of cognition. Zipf’s law may be related with language redundancy necessary for the language understanding process. This process can be modeled based on information compression performed by a self-organizing neural-computation structure at two levels. At the first level, a feature extraction is done in a parsing process of a natural conventional language, and the result is a linguistic map which acts as input for the second level of compression. There, a purely semantic map is formed which is independent of any conventional language, accounting in this way the universality of thinking and reasoning process.

Keywords: Cognitive modeling, language of thought, Zipf’s law, statistical linguistics.

1 Introduction

In a classical approach, the interdependence of language and thought has taken the shape of a theory called The Language of Thought Hypothesis (LOTH) [1]. However, because there is not yet a clear conception of thinking and language, different claims about their relation are expressed. Accepting the premise that language is related with the reasoning process, the main problem in any cognitive model is to account the process by which meaning is conveyed into sentences and the cognition which culminates in the intuitive perception of truth. The purpose of this article is an approach towards the developing of a cognition model starting from the premise of the validity of LOTH.

The second section makes a short presentation of the LOTH thesis. Section 3 shows that because any formal language is incomplete LOTH cannot be postulated in a similar fashion as a conventional language. Section 4 points to the possibility of using the Zipf's law in identification at the mind level of a deeper causal law responsible for language generation and understanding. Section 5 shows the possibility of modeling language understanding in a two-level hierarchy based on information compression. The last section draws the conclusions and points to further researches.

2 Language of Thought Hypothesis

LOTH appeared as an empirical thesis about the nature of thought and thinking. It is postulated that thought and thinking are performed in a mental language. However, since this idea of a mental level is itself a controversial one, i.e. there isn't yet a clear conception of mind accepted by scientific community, the language of thought is supposed to take place as a symbolic physical system realized in the brain of some appropriate organisms. Another postulate in this approach is that thoughts are defined as "propositional attitudes" [2, 3]. Propositional attitudes are the thoughts described in a language by sentences of the form: ***S A*s that *P***, where *S* refers to the subject of the 'attitude', *P* is the proposition that is the object of the attitude, and *A* stands for attitude verbs such as 'think', 'believe', 'hope', 'desire', etc. We see that from the way thoughts are defined, LOTH is based on the "Representational Theory of Mind" (RTM) [4]. The intimate relation language-thought is stressed by assuming that: "*S A*s that *P*" describes a mental process S - Ψ - $R(P)$, where there is a psychological relation Ψ between the subject S and a mental representation $R(P)$. $R(P)$ means that P. Mental representations are functionally describable entities that manifest at some suitable physical level, and the relation Ψ is meant to be understood as a computational/functional relation. So, the mental representations are a language with syntax and semantics.

We will see in the next section that conceiving mental representations similar with a formal conventional language is inadequate for accounting cognition due to the inherent limitations of any formal system.

3 Limitations of Language

Accepting the interdependence of language and thought leads us to the conclusion that all knowledge must culminate in verbal knowledge. Our knowledge of reality is shaped by the language we use. However, language has its own limitations in representing reality. A language is by its own nature a formal system, and any formal system is incomplete.

One way of seeing the incompleteness is from Gödel's theorem viewpoint, which makes a clear distinction between truth and provability. The meaning of truth and falsity exists in one's consciousness, but one is constrained to reason in some given context or language. Reasoning and truth recognition are two different things. In other words, this incompleteness, this Gödel phenomenon, points to the distinction between language and thought. The conclusion is that we have to give up the idea that everything, in order to be accepted, must be proven in a formal way or using a language. New conjectures and axioms should be chosen because of their usefulness and of large amounts of evidence suggesting that they are correct. But in order to decide upon the mathematical truth of such statements as Gödel propositions we shall need to employ insights or information from outside the system. A formal system cannot exceed its informational complexity, or its initial semantic content.

Next, we analyze the capacity of language to represent reality. Although LOTH posits an intimate relationship of language with reality, this intimacy seems to be superficial, at least with the kind of conventional language we use to communicate. This is a paradoxical situation, since we cannot imagine a world beyond language. Language is composed of words. The capacity of language to represent reality is given by the capacity of words to individually evoke an object. A word is a mere indicator of an object, it only reveals the object. Yet, the characteristics of that object, i.e. its form, size, color, attributes, are understood from our repeated observation and usage rather than from words. An object in its totality is never understood from the word which is used to denote it. The word is only a name given to an object. Yet the objects we perceive are not names. We have to make a distinction between the word which is an abstraction, existing in our minds, and the real object denoted by that word. Of course, there is also another distinction between the real object and the perception of this object by the senses. This three-level distinction suggests a possibility of a multilevel cognitive model, involving three different domains. One domain is the information domain. Another domain is the physical world, which we may know indirectly, through our senses. And, finally, the domain that we know directly is the domain containing mental images, and where sensorial sensations and thoughts are formed. The main difficulty in accepting this philosophical picture is to conceive the way in which the abstract informal domain, has influence upon the physical domain. Still, our common sense experience suggests that there is a close and real relationship between the informal domain and both the physical and mental one. As we shall see later, such a multilevel model can be developed by extending LOTH.

In any conventional language, every word is an arbitrary name given to an object. Thus, we may represent the reality through language by assigning names to things. Starting from Berry paradox, Rucker pointed out in that the concept of "nameability" is itself unnameable, i.e. that there can never be a reasonable short description (using a language) of how to understand language [5]. In this paradox we deal with naming numbers, instead of naming things, but ultimately the naming process involved is similar. In short, the Berry paradox shows that with *n* bits or words or whatever symbols we cannot explain or name what we understand by a name of *n* bits or words long. In other words, this proves the incompleteness of language to describe how the meanings are carried by words. There is no way to describe in a finite way how a series of words are transformed into thoughts or names into numbers.

Language is not only a mismatch for reality, it also misrepresents reality. We may arrange many combinations of words that represent something that doesn't exist in reality. The reality with which language is directly connected exists only in the mind. Thus whatever can be expressed by words belongs to mental reality. Language is not at all connected with the external world. It is connected with the mental reality, and also is incomplete in representing its own mental reality. In this context of language limitations, postulating LOTH as having similar syntactic/semantic form as a conventional language appears fallacious. Therefore, we need to identify other causal laws accounting for the structure of the mind. More, if language is related with thought, language properties would have to have a causal root in the functioning of the mind. This issue is controversial and is discussed next in relation with Zipf's law.

4 Zipf's Law

During the last century several scaling phenomena have been empirically discovered. These statistical laws are related somehow with living societies. Scientists are preoccupied to prove mathematically if these laws are universal and find out the causal factors behind the observed phenomena. One of the first concepts proposed to account for the hyperbolic type distribution found by Zipf in natural language texts was related with the principle of least effort, and seen somehow as a property of mind. However, this principle was not widely accepted and at present the scaling phenomena have not much relevance for cognitive science. Still, some authors try to find new evidence which may reveal an underlying cause of these phenomena and go beyond the laws of probability. The purpose of this section is to explore the possibility of relating scaling phenomena to cognitive modeling. We discuss in particular the Zipf's law and try to identify some deep causes that might be useful in accounting the cognitive function.

Analyzing the distribution of words in English texts, Zipf found a regular statistical pattern [6]. The most common word in English is 'the' and appears approximately twice as often as the second most common word, three times as often as the third most common, ten times as often as the tenth most common, and so on. The law is consistent also with the nature of communication, according to which the most common words tend to be short and appear often. What Zipf discovered is a power law. This means that small occurrences are extremely common, whereas large instances are extremely rare. Another example where the same law applies is the populations of cities. The ranked incomes also follow the same law. Zipf's law can be applied when the observed objects have a property (length, size, etc.) which is modeled by an exponential distribution that places restrictions on how often larger values can occur. The underlying principle is that efficiency, competition, or attention with regards to resources or information tends to result in Zipf's law holding to the ranking of objects.

Finding an explanation of Zipf's law in spite of some attempts to consider it irrelevant is still of much interest. We have only two options: (1) either we can assume that the law reflects some universal property of human mind, or (2) we can assume that it represents some necessary consequence of the laws of probabilities. The first approach is synthetic and was chosen by Zipf himself who proposed the validity of a principle of least effort that would explain the apparent equilibrium between uniformity and diversity in the common use of words. The second one is analytic, the law being viewed as a consequence of regarding a message source as a stochastic process.

The principle of economy Zipf referred to was not a new concept in science, and is natural for our common sense experience. The same principle regarding information processing in visual perception was previously considered. However, this concept was abandoned in favor of a search for a probabilistic explanation. Mandelbrot was the first who has pointed out that Zipf's law can be observed in "monkey typing" texts [7]. Miller also suggested that phenomena described by Zipf's law could be caused by a simple random process for creating words and the boundaries between words in natural languages, without being the need for appealing to least effort, least cost, or maximal information principles [8]. However, it's hard to conceive how the process of letter sequencing in forming a language and expressing meanings is nothing more than a random process that have a statistical explanation. A. Tsonis, Schultz, and P. Tsonis argue against random texts relevance to natural language for the purpose of proving Zipf's law [9]. They claim that in contrast with natural languages, in random texts all combinations of letters are considered as possible words, and the frequency of a word's occurrence is a function of its length. Very short and very long words are expected to be the least probable. They also suggest that the deviation from the Zipf's law or the under-representation observed for the higher-ranked words is not accidental, but is due to languages evolution and structure. In contrast, Li comments the above claims defending the statistical point of view [10]. In his opinion, the very short words are the most probable in random texts, and the Zipf's law manifests not only in unbiased random texts but also in biased texts as shown in [11, 12], when a monkey not necessarily types all alphabets with equal probability. Thus, by reducing the probability of some symbols not all combinations of letters can appear, and it is possible that a longer word will rank higher than a shorter one because it contains the more frequent symbols. Therefore, these differences put in discussion in [9] between random texts and natural languages seem to be irrelevant to the proof of Zipf's law in random texts. Zipf's law still holds even without an intentionally least effort principle involved in human communication. In conclusion, Li suggests that because the Zipf's law can be proved to exist in random texts, a "cost cutting"-like process cannot be identified as a deep causal law in natural language. In a more recent paper, Ferrer i Cancho and Sole defend the original least effort principle invoked by Zipf, [13]. They argue that Zipf's law is the result of the nontrivial arrangement of word–concept associations adopted for complying with hearer and speaker needs. In other words, Zipf's law is not a meaningless feature but a necessary condition for an optimum symbolic communication.

If the principle of least effort would prove to be right, and hence would be validated as being a property of mind, this would be of great importance for cognitive science, because at least we may identify a structural property at this level. Let's review what this principle says. According to Zipf, the development of language involves reaching a certain vocabulary balance as a result of two opposing forces, the force of *unification* and the force of *diversification*. The first force corresponds to the principle of least effort from the transmitter's point of view. This force tends to reduce the vocabulary. The second force has an opposite effect and is related with the receiver interest to associate meaning to speech, as much as possible. This model involving a two-person game, is somehow related with the role coding plays in classical theory of communication. The purpose of channel coding is to encode the source information in an effective manner so that a minimal amount of errors will occur when it is decoded. One effect of this is that channel coding employs redundancy to accomplish this. Source coding or information compression is used to reduce the number of bits transmitted from a source. The result is a lower redundancy in the transmitted information. It is very interesting to note that channel and source coding have opposite effects, one increases the size of the message and the other one decreases it. In the case of channel coding the redundancy is useful at the receiver for error detection and correction. On the other hand, the redundancy in the source message is dependent on the message to be sent and is not structured and useless for the receiver. We can see that the economical and efficiency principle of transmitting information is at the base of Zipf's approach. However, this has to be proven as being a causal principle at the mind level.

From the cognitive science point of view, the premise we suggest to consider in this discussion is that the language generation as comprising the semantic information content of the transmitter must have a causal law beyond the statistical law behavior. If natural language is of the same nature as random texts then the semantic association with the syllables of words would be completely formless, and no semantic structure could be identified in language. Words convey meaning in a correspondence to their syllables structure. A word is nothing more than the phonemes that are found in it, but these very phonemes cause the understanding of the word's meaning. The process of comprehension resides in the accumulation in the hearer's (or reader's) mind of the syllables impressions as they are sequenced in time. These syllables make invariant words but also inflectable word-bases and inflections. A word-base or an invariant word contributes its meaning to the sentence meaning in an independent way, while an inflection contributes its meaning in a dependent way. Therefore we have to understand the word as a sound/symbol sequence which has to support syllables inflections.

In this context of word dependency of the syllables structure, we studied the word frequencies of a random word generator based on the analysis of the frequency pattern of pairs of letters in English word-list databases. Using Markov chains the algorithm starts by printing any *n* consecutive letters in the text. Then at every step it searches for any random occurrence in the original text of the last *n* letters already printed and then prints the next letter. Some letters are more likely than others to occur after a given pair of letters. In this way the words generated are devoid of any semantic content but nonetheless are syntactically correct, i.e. pronounceable. We want to prove the influence of higher semantic level in language structuring. We used two random word generators [14, 15]. Three experiments, labeled (i), (ii), using the first generator, and (iii), using the second generator, have been performed as presented in Table 1. The words generated vary in length from 2 to 24 symbols. The most frequent words occupying the first ranks are 2-letter, 3-letter and 4-letter length for all cases. It can be seen that the word frequency has a decreasing trend with the rank but it cannot be associated with a power law pattern due to its very low significance in the total amount of words. The first ranked words have a weight of only 0.6% - 1.5% of the total number of generated words, which is much too less for being relevant for a statistic pattern.

Table 1
Word profile statistics for 3 experiments of randomly generated words.

Word frequency			No. of words			Percentage word count		
i	ii	iii	I	ii	iii	i	ii	iii
10	15	6	1	1	1	100	100	100
5	4	5	1	1	2	99.03	98.58	99.43
4	3	3	1	3	1	98.55	98.20	98.49
3	2	2	3	36	21	98.17	97.38	98.20
2	1	1	15	960	999	97.3	90.56	94.25
1	-	-	982	-	-	94.42	-	-

For the sake of comparison, the word frequency versus rank has been computed for a natural language text having a close length. We chose *The Library of Babel* by Luis Borges, which contains around 1020 words. The word frequency and the word count percentage are given in Table 2 for the first 20 ranks which contain only one word, excepting the ranks 18 and 20 having 2 words. The first ranked word is “the” and appears 223 times, which means 21.84% of the total number of words. The list of the most 20 frequent words is given in Table 3.

Table 2
Word statistics for natural language text (*The Library of Babel* by Luis Borges).

word freq.	% word count	word freq.	% word count	word freq.	% word count	Word freq.	% word count
223	100	59	80.9	30	72.9	22	68.4
130	92.3	56	78.9	28	71.9	21	67.6
75	87.9	51	77	27	71	20	66.9
66	85.3	36	75.2	25	70	19	65.5
62	83	31	74	24	69.2	18	64.9

Table 3
The first 20 ranked words in *The Library of Babel*.

rank	word	rank	Word	rank	word	rank	Word
1	the	6	is	11	it	16	have
2	of	7	that	12	library	17	one
3	and	8	to	13	this	18	books
4	a	9	which	14	not	19	are
5	in	10	I	15	be	20	all

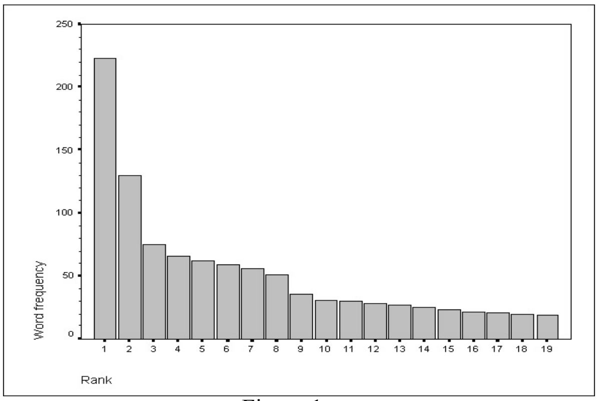

Figure 1
Word frequencies versus rank for natural language text.

The frequency versus rank is depicted in Figure 1. We can see that a typical hyperbolic pattern emerged, which indeed proves the Zipf’s-like law tendency as expected for natural language texts.

In the next section we propose a two-level hierarchical model of language based on information compression. At the first level, language and speech manifest externally. The meaning being encoded into letters/phonemes, in order to be understood language needs to be redundant. The employ of natural language in a two-person like game involves redundancy and hence information compression. This redundancy as we might expect is employed in helping the receiver to correct errors and compensate for noise. In our example, we can see from Table 3 that the most frequent words are less informative. The first word bearing a semantic content related to the narration's context is "library" and occupies only the 12th position. The word 'books" is the next semantic important one and comes in the 18th position. On the other hand, there is another property of language. Every word in a natural language can be regarded as a code or label for its meaning. The meaning is a relatively complex body of knowledge, and the word acts like a tag in pointing to that knowledge. Therefore, even if natural language text manifests redundancy similar with channel coding, the source coding is performed also by coding the information into the words' symbolism. Therefore, sentences are highly coded and compressed representations of speaker's information source.

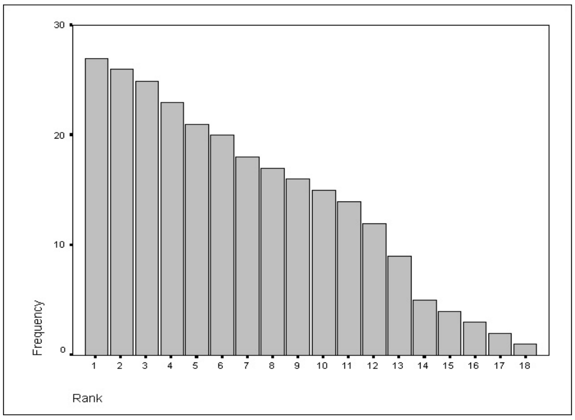

Figure 2
Word frequency versus rank for the compressed text.

A compressed sequence manifests random properties, and if the power law would manifest as a statistic property of random texts this would have to be visible after compression also. The same text was compressed using Lempel-Ziv algorithm, and the frequency variable distribution is shown in Figure 2. The distribution looks much more linear than the previous one for the uncompressed text, without the power-like law tendency.

The above results show that Zipf's law doesn't manifest at a higher level of semantic cognition where language appears compressed. Therefore, Zipf's law can be rooted in a language structuring process of coding, which adds redundancy necessary for language understanding. This would take place at the second level in the model we propose. Language generation implies redundancy according to the observed power-like law, and language understanding implies eliminating redundancy by compression.

5 Two-Level Language Hierarchy

From the viewpoint of algorithmic complexity, the most fundamental property of brain has to be the enormous interdependence between its components. This means that it is much simpler to view it as a whole than as the sum of its parts. If both, the whole and parts are equally complex then the parts are independent, i.e. they do not interact. This suggests the idea of information compression process that might take place at some higher level of the cognitive architecture. The process of inductive reasoning itself suggests also this idea. It involves the mental merge of the repeating instances of each pattern in our experience, a feature extraction from the perceived data. If there is no pattern, there is no understanding. That is the reason why a random pattern seems to be valueless: because our failure in the attempt of extracting redundancy or compressing it. There is also other evidence from psychology, biology and computer science [16, 17], showing that information compression is used as technique in cognitive processes.

All these arguments point to the conclusion that information compression is an essential feature of information systems, being intimately related to the principle of inductive reasoning, which itself provides a foundation for any cognitive system dealing with the storage and processing of information. We suggest that LOTH may be applied following the principle of information compression. Several authors developed symbolic methods for parsing natural language based on information compression [18, 19]. The basic idea is searching for redundancy which can be understood as a search for patterns which match each other. Like any other search problem because of its exponential nature it has to deal with finding an optimum strategy in terms of search costs. Therefore we can expect to be normal to circumscribe the search or to guide the search by some measure of redundancy or both in the process of inductive inference. However, a symbolic approach involving searching can hardly deal with real situations. The idea in our approach is to use a self-organizing neural-computation approach of data compression. A self-organizing network can be trained to perform a mapping F: $X^n \rightarrow Y^m$, where $m < n$. The input information or the training pattern is redundant. The network is capable of performing a feature extraction or a dimensionality reduction of input data. This kind of neural-computation approach seems very suitable for performing data compression as we need in our model of language understanding. Training is unsupervised and is usually based on a form of global competition between neurons. We are interested in self-organizing networks that cluster or categorize the input data. The architecture should be developed on two levels of language processing. On the first level, language is parsed according to the conventional grammar rules of whatever language is analyzed. Here, a remark is necessary: obviously, there are certain differences between spoken and written language processing, especially at the early stages of processing. However, it is not our purpose to deal with these differences. The issues that we shall explore in this article are common to both spoken and written languages. The neural network at the first level performs a mapping F_{CL}: $X^n \rightarrow Y^2$. The map represents a twodimensional projection of the distance relations between the patterns in the original multidimensional pattern space of a conventional language. The learning rule follows the classical competitive learning. The input multidimensional space is formed of linguistic categories as in the classical parsing approach (NP, VP, PP, etc.). The output may be in principle a phonotopic map or a syntactic-semantic map. The function of the first mapping level is to compress language features or regularities of any conventional language to an intermediate level of understanding. At this level, the semantic properties of language are only partially represented.

One difficulty here is to define a suitable metrics for the words of a language. The distance function on the input space of patterns has to be in harmony with the choice of appropriate input vectors $\boldsymbol{x}$. The different clusterings of the input data depend on how this choice of metrics is done. Another problem is that at this purely conventional language interpretation, the meaning of a word is disassociated from its encoding into letters and/or sounds. Also, the words denoting similar objects should be mapped with the same topology. Therefore, a metric relation should be defined between these words. A classical solution is due to Ritter and Kohonen, and developed by Honkela [20, 21]. The conclusion is that even in a limited context, a meaningful semantic map can be produced. The neural-computation approach proved to be successful in dealing with expressions that are of symbolic nature, and could display semantic relationships among them.

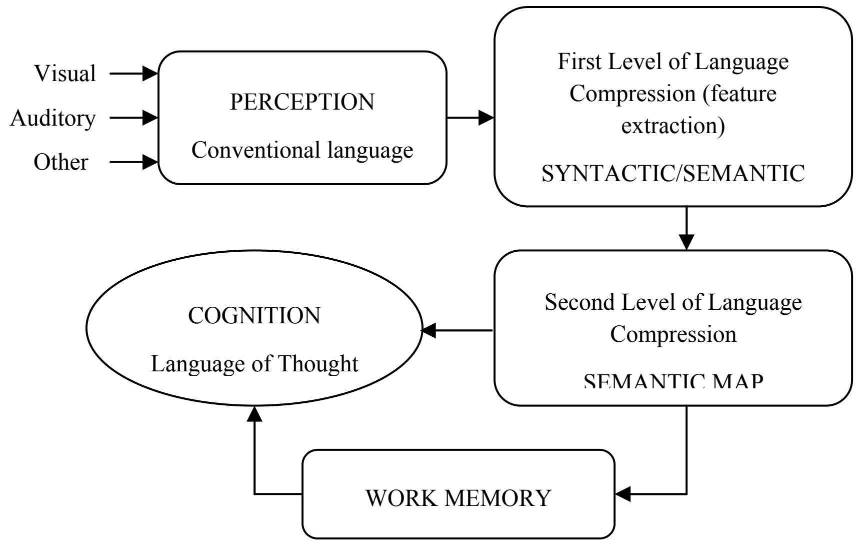

Figure 3
Model of cognition based on language compression

Next, we need a second neural network that will accept as input the syntactic-semantic maps obtained on the first level. The result will be a purely semantic language independent map, denoting only higher concepts or abstractions. These form a language of thought embracing the whole domain of knowledge, and used for thinking. This language is independent of the symbolic conventional language used in communication, and does not contain distinguishable symbolic parts, as an ordinary communicative language. In Fig. 3 it is presented a general structure of a possible cognitive model. The existence of a second level of language is supported also by psychological experiments in which the subjects are asked after reading a certain text to reproduce the exact words and the phrase they read. The subjects remembered the whole idea of what they read in most of the cases, but rarely they could chose correctly the exact sentence they read. Recent experiments suggest higher memory retention for meaningful words and pictures versus meaningless ones [22].

The above proposed model goes beyond the premise that every language is conventional and temporal, and hence the relationship between word and meaning has to be purely arbitrary. Such a view would lead us to the fallacy of infinite regress. Therefore, we have to postulate that if language of thought would exist, this has to be universal and manifested in a completely compressed form. Therefore, it is not expected that any distinctive lexical parts would be discernible. This is accomplished in the proposed model by the second level of language compression. Here, the semantic map constitutes the mental representation of concepts and abstractions, the semantic patterns of objects, actions, and attributes. The semantic map represents in fact a metalanguage, i.e. a language in which one can describe both objects in the world and sentences in the object language. By the existence of this level we can account solving paradoxes such as the liar paradox: (S) “This very sentence id false.” The truth of such a sentence belongs to a language of a higher order, in our model the compressed language of thought. In the proposed model, at the perceptive level, language and speech manifest externally, existing a difference between the meaning of a word and its encoding into letters and/or sounds. Ultimately, at the highest level of compression there is no necessity of sequence in language. Meaning appears in its entirety as a semantic map. By these two levels of language we also can explain the duality of rationalistic and intuitive thinking. The communicative language has to comprise distinctive sound/symbol units through which one can express rationalistic thoughts. On the other hand, at a higher level of understanding process, the inner state of cognition is ineffable. In this state of conscious awareness the meaning is perceived in its entirety as a whole, without the need of mediation through distinctive lexical parts.

Conclusions

We started from the premise that thinking and speaking are interrelated and there is a language of thought for representing knowledge. At one hand, we use a communicative language, through which we can express our thoughts, but in the same time we are aware that much of our knowledge is represented in nonverbal form. This indicates that knowledge is not ultimately represented symbolically by the aid of words, but by a sort of noncommunicative purely semantic representation. Only, when we communicate something, a conventional set of symbols are used in forming a language. Even if there are so many different languages (and many more can be in principle invented) there is no evidence that the speakers of those languages think about the world in different ways. Therefore, we propose a two-level hierarchical model of language based on information compression. At the lowest level, language is used conventionally in a symbolic form, for communicating ideas and concepts that are reasoned at a higher level in a universal language of thought manifested in a completely compressed form. Language generation implies redundancy according to the observed power-like law, and language understanding implies eliminating redundancy by compression in two levels as presented above.

References

[1] A. J. Fodor: The Language of Thought. New York: Thomas Y. Crowell Co., 1975

[2] A. J. Fodor: Propositional Attitudes. In Representations: Philosophical Essays on the Foundations of Cognitive Science, J. A. Fodor, Cambridge, Massachusetts: MIT Press, 1981

[3] M. Aydede: Connectionism and Language of Thought, Tech. Rep., Stanford, CSLI-95-195

[4] A. J. Fodor: Psychosemantics: The Problem of Meaning in the Philosophy of Mind. Cambridge, Massachusetts: MIT Press, 1987

[5] R. Rucker: Infinity and the Mind. Princeton University Press, New Jersey, 1995

[6] George K. Zipf: Human Behaviour and the Principle of Least-Effort. Addison-Wesley, Cambridge MA, 1949

[7] B. Mandelbrot: Linguistique Statistique Macroscopique. In L. Apostel, B. Mandelbrot and A. Morf, Logique, Language et Theorie de l'Information, Paris: Presses Universitaires de France, 1957

[8] G. Miller: Some effects of intermittent silence. Am. J. of Psychology. 70: pp. 311-314, 1957

[9] A. A. Tsonis, C. Schultz, and P.A. Tsonis: Zipf's law and the structure and evolution of languages. Complexity. 2(5): pp. 12-13, 1997

[10] W. Li: Letters to the Editor. Complexity, 3(5): 9-10, 1998

[11] W. Li: Random texts exhibit Zipf's-law-like word frequency distribution. IEEE Transactions on Information Theory, 38(6), pp.1842-1845, 1992

[12] R. Perline: Zipf's law, the central limit theorem, and the random division of the unit interval. Physical Review E, 54(1), pp. 220-223, 1996

[13] Ramon Ferrer i Cancho, Ricard V. Sole: Least effort and the origins of scaling in human language. Proc. Natl. Acad. Sci. USA, vol. 100, no. 3, 2003, pp. 788-791

[14] R. Lawrence: Random Word Generator. http://www.fourteenminutes.com/fun/words/

[15] A. Chase: A Generator of Perfectly Cromulent Words. http://andy.greyledge.net/misc/cromulac/

[16] J. G. Wolff: Computing, Cognition and Information Compression. El. Journal on Theory and Review in Psychology, 1997 (http://gemstate.net/susan/Wolff.htm)

[17] Ming Li and Paul Vitanyi: An introduction to Kolmogorov complexity and its applications. Revised and expanded second edition, Springer-Verlag, New York, 1997

[18] J. A. Storer: Data Compression: Methods and Theory. Computer Science Press, Rockville, Maryland, 1988

[19] J. G. Wolff: Parsing as information compression by multiple alignment, unification and search: SP52. SEECS Report, 1998

[20] H. Ritter, and T. Kohonen: Self-Organizing Semantic Maps. Biolog. Cybern. 61: 241-254, 1989

[21] T. Honkela: Self-Organizing Maps in Natural Language Processing. PhD thesis, Helsinki University of Technology, 1997

[22] E. Mu and D. F. Galleta. The Influence of the Meaning of Pictures and Words on Web Page Recognition Performance. Proc. of the 36th Hawaii Intern. Conference on System Sciences (HICSS'03)