Chapter V

Symbols: Integrated Cognition and Language

Leonid I. Perlovsky
Air Force Research Lab, USA

Abstract

What is the nature of symbols? This word is used for traffic signs, mathematical notations, and motivationally loaded cultural objects that may inspire war and peace. This chapter explains relationships among symbols, cognition, and language. Symbols are explained as processes in the mind involving cognition and language. Relationships between cognition and language were a mystery until recently. Linguists often considered language as relationships among words and other linguistic entities, separately from its relationships to the world. Mechanisms of language in the mind and brain were considered separate and different from thinking and cognition. Neural mechanisms integrating language and cognition are unknown. Yet, language and cognition are intertwined in evolution, ontogenesis, learning, and everyday usage; therefore, a unified understanding of working of the mind is essential. A mathematical description of such unifying mechanisms is the subject of this chapter. We discuss relationships among computational intelligence, known mechanisms of the mind, semiotics, and computational linguistics, and describe a process integrating language and cognition. Mathematical mechanisms of concepts, emotions, and instincts are described as a part of information processing in the mind and related to perception and cognition processes in which an event is understood as a concept. Development of such mathematical theories in *the past often encountered difficulties of a fundamental nature manifested as combinatorial complexity. Here, combinatorial complexity is related to logic underlying algorithms, and a new type of logic is introduced—dynamic fuzzy logic—which overcomes past limitations. This new type of logic is related to emotional signals in the brain and combines mechanisms of emotions and concepts. The mathematical mechanism of dynamic logic is applicable to both language and cognition, unifying these two abilities and playing an important role in language acquisition as well as cognitive ontogenesis. The mathematical description of thought processes is related to semiotic notions of signs and symbols.*

Symbols in Computational Intelligence and Linguistics

Symbol is the most misused word in our culture (Deacon, 1998). We use this word in trivial cases referring to traffic signs and in the most profound cases of cultural and religious symbols. Charles Peirce (1897, 1903) considered symbols to be a particular type of signs. He concentrated on the process of sign interpretation, which he conceived as a triadic relationship of sign, object, and interpretant. Interpretant is similar to what we call today a representation of the object in the mind. However, this emphasis on interpretation was lost in the following generation of scientists.

In the development of scientific understanding of symbols and semiotics, the two functions—understanding language and understanding world—often have been perceived as identical. This tendency was strengthened by considering logic to be the mechanism of both language and cognition. According to Bertrand Russell (1919), language is equivalent to axiomatic logic, a word-name "merely to indicate what we are speaking about; [it] is no part of the fact asserted … it is merely part of the symbolism by which we express our thought." David Hilbert (1928) was sure that his logical theory also describes mechanisms of the mind: "The fundamental idea of my proof theory is none other than to describe the activity of our understanding, to make a protocol of the rules according to which our thinking actually proceeds."

Logical positivism centered on "the elimination of metaphysics through the logical analysis of language," according to Rudolf Carnap (1928) logic, was sufficient for the analysis of language. This belief in logic has deep psychological roots related to the functioning of the human mind. A major part of any perception and cognition process is not accessible to consciousness directly. We are conscious about the final states of these processes, which are perceived by our minds as concepts approximately obeying formal logic.

Similar understanding of relationships among symbol, language, logic, and mind can be traced in semiotics of Ferdinand Saussure (1916) and in structuralism. A simplistic idea that words are labels for objects falls apart as soon as we consider words for abstract ideas, say, *rational*. Saussure (1916) tried to resolve this problem by saying that "the linguistic sign does not unite a thing and a name, but a concept and a sound image." Here, the real world is taking a back seat; both aspects of the sign exist in the mind. Structuralism was derived later from Saussurean linguistics. It emphasized "concept" as a part of language and pushed semiotics further from the real world, further from the mind, toward relationships among words. This movement away from the world toward words in semiotics was inspired by a similar movement in logic toward axiomatic meanings. Formal logicians emphasized that the foundation of our knowledge is limited to abstract mathematical objects and axioms that they obey. Relationships between mathematical objects and the world are arbitrary. Similarly, Saussure emphasized the arbitrariness of the sign; relationships between words and objects in the world are arbitrary conventions.

This idea later evolved into arbitrariness of communication codes in general. Since communication codes contain cultural values, some concluded that cultural values are arbitrary. "There may be an objective, empiricist reality out there, but there is no universal, objective way of perceiving and making sense of it. What passes for reality in any culture is the product of the culture's codes, so 'reality' is always already encoded, it is never 'raw'" (Fiske, 1987). This circle of ideas served as a platform for Jacques Derrida's (1978) attacks on structuralism. Since any statement is based on some cultural structures and values, it can be dismissed as having no objective validity, as arbitrary or as local. This became the essence of deconstruction and postmodernism. This reasoning can be applied to deconstruction itself, so the deconstruction is as if self-annihilated. The self-annihilation is not new to logicians; it is just a particular case of Gödelian (1929–1936) proof that logic is not logical. Derrida (1976) understood deconstruction as a question. In this chapter, we attempt to give an answer to this question—How is it possible to have anything of truth and value?—how our mind constructs symbols, which have psychological values and are not reducible to arbitrary signs.

An idea that language and cognition were not one and the same and that logic was not a fundamental mechanism of the mind slowly downed in the contemporary science after Kurt Gödel (1929–1936) proved inconsistency of logic. Logic turned out to be not the queen of mathematics, as was previously believed. Yet, early computational theories of the mind developed since the 1950s heavily relied on logic. In 1957, Noam Chomsky proposed that language ability was based on inborn mechanisms of language faculty, which was a system of logical rules. Similarly, logical rules were the basis for the Artificial Intelligence developed from the 1960s through the 1980s (see Minsky, 1968). During that time, much evidence was accumulated, indicating that computer systems based on logic were not successful in simulating human abilities for language or cognition. New mathematical methods not based on logic were emerging, which included Stephen Grossberg's (1982) neural networks, Chomsky's (1981) principles and parameters, methods based on fuzzy logic (1965), and the author's modeling field theory and dynamic logic (Perlovsky, 1987, 2001).

Linguists, following Chomsky, emphasized that cognition and language abilities are different: they are located in different brain areas and might have emerged along separate paths in evolution (Pinker, 2000). Most importantly, cognition is about objects and situations in the surrounding world, whereas mechanisms of acquiring and using language identified in cognitive linguistics are about language, not about the world. This direction, emphasizing innateness of language abilities, was called *nativist linguistics*. Its interests concentrated on the internal mind mechanisms enabling learning of language. Chomsky emphasized the importance of syntax; he considered its mechanisms to be relatively isolated from the rest of the mind and brain. Relations of language to the outside world, to perception, cognition, and meaning were considered peripheral to the study of language.

An opposite direction—cognitive linguistics—appeared in the 1970s. Several linguists, including George Lakoff (1987), Ronald Langacker (1987), and Leonard Talmy (1988), emphasized cognitive principles and organization of the mind and considered meaning as central to language. This appreciation that language cannot be understood separately from thinking about the world was shared by researchers in computational semiotics (Perlovsky, 2002, 2004; Rieger, 2002) and in evolutionary linguistics (Brighton et al., 2005; Christiansen & Kirby, 2003; Hurford, Studdert-Kennedy, & Knight, 1998).

Today, little is known about neural mechanisms combining language with thinking or their locations in the brain (Deacon, 1998; Lieberman, 2000). In the following sections, we briefly review research in cognition and linguistics. We analyze fundamental mathematical difficulties faced by researchers in these fields. Then, we describe new approaches overcoming these difficulties and discuss a mathematical theory of symbols integrating cognition and language.

Combinatorial Complexity of the Mind Theories

Understanding signals coming from sensory organs involves associating subsets of signals corresponding to particular objects with internal representations of these objects. This leads to recognition of the objects and activates internal brain signals leading to mental and behavioral responses, which constitute the understanding of the meaning (of the objects).

Mathematical descriptions of this association-recognition-understanding process have not been easy to develop; a number of difficulties have been encountered during the past 50 years. These difficulties have been summarized under the notion of combinatorial complexity (CC) (Perlovsky, 1998a). The problem was first identified in pattern recognition and classification problems in the 1960s and was named "the curse of dimensionality" (Bellman, 1961). At first, it seemed that self-learning statistical pattern recognition algorithms could learn to solve any problem, only if sufficient amount of training data were provided. However, after decades of research, it became clear that adaptive statistical pattern recognition and neural network algorithms designed for self learning often encountered CC of learning requirements: the required examples had to account for all possible variations of an object in all possible geometric positions and in combinations with other objects, sources of light, and so forth, leading to astronomical (and worse) numbers of required examples (Perlovsky, 2001).

By the end of the 1960s, a different paradigm became popular: logic-rule systems were proposed to solve the problem of learning complexity. An initial idea was that rules would capture the required knowledge and eliminate a need for learning. The first Chomsky (1972) ideas concerning mechanisms of language grammar that were related to deep structure also were based on a similar idea of logical rules. Rule systems work well when all aspects of the problem can be predetermined. However, rule systems and expert systems in the presence of unexpected variability encountered CC of rules: more and more detailed subrules and sub-subrules, one contingent on another, had to be specified.

In the 1980s, model-based systems became popular, which were proposed to combine advantages of adaptivity and rules by utilizing adaptive models (Bonnisone, Henrion, Kanal, & Lemmer, 1991; Nevatia & Binford, 1977; Perlovsky, 1987, 1988, 1991). Existing knowledge was to be encapsulated in models, and unknown aspects of concrete situations were to be described by adaptive parameters. Along similar lines were principles and parameters ideas of Chomsky (1981). Model-based systems encountered computational CC (N and NP complete algorithms). The reason was that considered algorithms had to evaluate multiple combinations of elements of data and rules (models, principles). CC is prohibitive because the number of combinations is very large; for example, consider 100 elements (not too large a number), which combinations had to be evaluated. The number of combinations of 100 elements is 100^{100}; this number is larger than all interactions among all elementary particles in life of the Universe. No computer (or mind) would ever be able to compute that many combinations. The CC became a ubiquitous feature of intelligent algorithms and, seemingly, a fundamental mathematical limitation.

Forms of the Mind, Logic, and CC

Logic serves as a foundation for many approaches to cognition, linguistics, and computational semiotics. Logic underlies most of computational algorithms. But its influence extends far beyond, affecting psychologists and linguists who do not use complex mathematical algorithms for modeling the mind. All of us operate under a more than 2,000-year-old influence of logic under a more or less conscious assumption that in the basis of the mind, there are mechanisms of logic. As discussed next, our minds are unconscious about its illogical foundations; we are mostly conscious about a small part about the mind's mechanisms, which are approximately logical. Our intuitions, therefore, are unconsciously affected by bias toward logic. When laboratory data drive our thinking away from logical mechanisms, it is difficult to overcome the logical bias.

However, relationships between logic and language have been a source of longstanding controversy. Aristotle assumed a close relationship between logic and language. He emphasized that logical statements should not be formulated too strictly and that language inherently contains the necessary degree of precision. Aristotle described the mechanism of the mind relating language, cognition, and the world as *forms*. Today, we call similar mechanisms internal representations or concepts in the mind. Aristotelian forms are similar to Plato's ideas with a marked distinction: forms are *dynamic*—their initial states, before learning, are different from their final states of concepts (IV BCE). Aristotle emphasized that initial states of forms (forms-as-potentialities) are not logical, but their final forms (forms-as-actualities) attained in the result of learning are logical. This important distinction was lost during millennia of philosophical arguments.

The founders of formal logic emphasized a contradiction between logic with its law of excluded third and language with its uncertainty. In the 19th century, George Boole and great logicians following him, including Gottlob Frege, Georg Cantor, David Hilbert, and Bertrand Russell, eliminated uncertainty of language from mathematics and founded formal mathematical logic based on the "law of excluded third." Hilbert developed an approach named "formalism," which rejected intuition as a matter of scientific investigation and formally defined scientific objects in terms of axioms or rules. In 1900, he formulated the famous Entscheidungsproblem to define a set of logical rules sufficient to prove all past and future mathematical theorems. This entailed formalization of the entire human thinking, including language. Formal logic ignored the dynamic nature of Aristotelian forms and rejected uncertainty of language. However, Hilbert's vision of formalism explaining mysteries of the human mind came to an end in the 1930s, when Gödel proved internal inconsistency of formal logic. This is a reason why theories of cognition, language, and computational semiotics based on formal logic are inherently flawed.

It turned out that combinatorial complexity of algorithms based on logic is a finite-system manifestation of Gödel's theory (Perlovsky, 1996a). According to the law of excluded third, every statement is either true or false and nothing in between. Therefore, algorithms based on formal logic have to evaluate every little variation in sensory signals or the mind's representations as a separate logical statement. A large number of combinations of these variations causes combinatorial complexity. The CC of learning requirements of pattern recognition algorithms and neural networks is related to logic; every training sample is treated as a logical statement, which results in CC. Multivalued logic and fuzzy logic were proposed to overcome limitations related to the law of excluded third (Kecman, 2001). Yet, the mathematics of multivalued logic is no different in principle from formal logic. Fuzzy logic encountered a difficulty related to the degree of fuzziness: if too much fuzziness is specified, then the solution does not achieve a needed accuracy; if too little, it becomes similar to formal logic. If logic is used to find the appropriate fuzziness by sorting through various degrees of fuzziness for every model at every processing step, then it results in CC. Dynamic logic discussed later overcomes CC by automatically choosing the appropriate degree of fuzziness for every model at every step. This dynamics can serve as a mathematical representation of the learning process of Aristotelian forms.

Nativist Linguistics vs. Cognitive Linguistics

During the first half of the 20th century, there was little appreciation in linguistics and psychology for complicated innate mechanisms of the mind. There was no mathematics adequate for findings of Sigmund Freud and Carl Jung. Logic dominated thinking of mathematicians and intuitions of psychologists and linguists. Within logic, there is not much difference between language and cognition (both are based on logical statements), and there is no difference between signs and symbols. Motivational aspects of symbols were not addressed, and a mixup of symbols and signs continued. In the second half of the 20th century, a variety of mathematical approaches were tried in order to explain perception and cognition. As computers and robotics gained importance in engineering, huge efforts were expended toward making computers smarter. Knowledge about the human mind was used to enhance computer intelligence. As discussed, every mathematical method faced combinatorial complexity and failed. Mathematical methods did not explain how the mind creates meaning. Mathematical approaches in linguistics paralleled those in perception and cognition

In the 1950s, Chomsky moved linguistics toward studies of innate mind mechanisms. Nativists used available mathematics of logical rules similar to artificial intelligence. In 1981, Chomsky proposed a new mathematical paradigm in linguistics, rules, and parameters. This was similar to model-based systems emerging in mathematical studies of cognition. It was influenced heavily by logical bias and, as discussed, faced CC. In 1995, Chomsky's minimalist program called for simplifying rule structure of the mind mechanism of language. It moved language closer to other mind mechanisms and closer to the meaning but stopped at an interface between language and meaning. Chomsky's linguistics still assumed that meanings appear independently from language, and a mixup of signs and symbols continued; motivational forces of symbols were ignored.

Cognitive linguistics emerging in the 1970s intended to address some of these limitations of the nativist approach. Cognitive linguists wanted to unify language with cognition and explain creation of meanings. They were looking for simpler innate structures than those postulated by nativists. These simpler structures would be sufficient, scientists thought, because they would combine language and meaning and combine innate structures with learning from experience (to a much larger extent than nativists postulated). Cognitive linguists gradually moved away from the heavy influence of logical bias of the previous structuralist thinking, which could be characterized by "the meaning of a word can be exhaustively decomposed into finite set of conditions … necessary and sufficient" Jackendoff (1983).

Lakoff (1988) emphasized that abstract concepts used by the mind for understanding the world have metaphorical structure. Metaphors were not just poetic tools but rather an important mechanism of the mind for creating new abstract meanings. Lakoff's analysis brought this cultural knowledge, advanced by Fyodor Dostoevsky and Friedrich Nietzsche, within the mainstream of science. There was still a big gap between Lakoff's analysis of metaphors on the one hand and neural and mathematical mechanisms on the other. The "metaphors we live by" is a metaphorical book (pun intended) in that it begs these questions: Who is that homunculus in the mind, interpreting the metaphorical theater of the mind? What are the mechanisms of metaphorical thinking?

In the works of Jackendoff (1983), Langacker (1988), Talmy (1988), and other cognitive linguists,[1] it was recognized that old divisions dominating linguistics were insufficient. Dichotomies of meanings (semantic-pragmatic) and dichotomies of hierarchical structures (superordinate-subordinate) were limiting scientific discourse and have to be overcome. Consider the following opinions on meaning creation:

In a hierarchical structure of meaning determination the superordinate concept is a necessary condition for the subordinate one. … COLOR is a necessary condition for determining the meaning or RED. (Jackendoff, 1983)

The base of predication is nothing more than … domains which the prediction actually invokes and requires. (Langacker, 1988)

These examples illustrate attempts to overcome old dichotomies and, at the same time, difficulties encountered along this path. Both examples are influenced by logical bias. Attempts to implement mathematical mechanisms assumed by these examples would lead to combinatorial complexity. To put it jovially, problems of meaning and hierarchy still reminded the old question about the chicken and the egg: what came first? If superordinate concepts come before subordinate ones, where do they come from? Are we born with the concept COLOR in our minds? If predictions invoke domains, where do domains come from? These complex questions with millennial pedigrees are answered mathematically in the following sections. Here, I give a brief psychological preview of the answer, informed by contemporary development in dynamic logic, neurobiology, and language evolution. Hierarchy and meaning emerge jointly with cognition and language. In processes of evolution and individual learning, superordinate concepts (COLOR) are vaguer, less specific, and less conscious than subordinate ones (RED). RED can be vividly perceived, but COLOR cannot be perceived. RED can be perceived by animals. But the concept COLOR can only emerge in the human mind due to joint operation of language and cognition.

Jackendoff (2002), in his recent research, concentrated on unifying language and cognition. He developed detailed models for such unification; however, his logical structures face combinatorial complexity. Lakoff and Johnson (1999) brought within the realm of linguistics an emphasis on embodiment of the mind. The implication that the philosophical tradition will have to be reassessed, however, seems exaggerated. Recent synthesis of computational, cognitive, neural, and philosophical theories of the mind demonstrated the opposite (Perlovsky, 2001). Plato, Aristotle, and Kant, even in specific details about the mind mechanisms, were closer to contemporary computational theories than the 20th-century philosophers and mathematicians developing logical formalism and positivism.

Talmy (2000) introduced a notion of open and closed classes of linguistic forms. Open class includes most words, that could be added to language as needed, say, by borrowing from other languages. Closed class includes most grammatical structures, which are fixed for generations and cannot be easily borrowed from other languages. This pointed to an important aspect of interaction between language and cognition. Forms of the closed class interact with cognitive concepts, which emerged over thousands of years of cultural and language evolution. Thus, for each individual mind and for entire generations, which operate within constraints of existing grammar, many cognitive concepts are predetermined. Talmy identified cognitive concepts affected by closed forms. These forms are more basic for cognition than words and unconsciously influence entire cultures, as suggested by Nietzsche (1876). Current research into mechanisms of language-cognition interaction revealed a profound impact of these forms beyond merely conceptual and identified their impact on emotional contents of languages and cultures (Perlovsky, 2006b).

A controversy between nativists and cognitivists does not imply that linguists doubt the importance of innate mechanisms or the importance of learning and using language. Humans are the only species endowed with language; therefore, some mechanisms have to be inborn. Equally, there is ample evidence that a child will not learn language if not exposed to it (and the exposure must come during a specific critical period, possibly between 2 and 7 years old). The controversy is about what exactly is innate and what kind of exposure is sufficient for learning. In *Rethinking Innateness: A Connectionist Perspective on Development*, Jeffrey Elman et al. (1996) demonstrated that many aspects of language acquisition can be explained within the framework of connectionist neural network. They demonstrated that detailed syntactic rules postulated by nativists are not necessary and that learning of complex syntactic patterns still can occur without previous exposure to exactly the same patterns.

Elman (1993) continued this discussion of connectionist use-based language acquisition vs. nativist rule-based acquisition. The main argument again is that the innate mechanisms can be given by connectionist architectures much simpler than logical rules. But what is "simpler"? Elman emphasizes the other side of the story. The connectionist neural network is not an off-the-shelf multilayer perceptron but rather an SNR neural network carefully designed for language acquisition (Elman, 1990). Moreover, SNR performs not a general language acquisition but rather a specific type of learning for which it was designed. Elman (1993) emphasized a hard learned lesson that we discussed in previously: "there is no … general purpose learning algorithm that works equally well across domains" (p. 1).

Does it mean that our mind uses a huge number of diverse algorithms for language and cognition à la Minsky (1988)? Or there are fundamental first principles of the mind organization (see discussion in Perlovsky, 2004). I'd note that there are no more than 300 genes determining differences between the human mind and the ape mind. The mechanisms for language and cognition cannot be too specific; our abilities are adaptive, and any child can learn any of the thousands of languages spoken on Earth. We, therefore, have reasons to believe in first principles of the mind. SNR neural network cannot be an example for such a general principle; according to analysis in previous sections, SNR will face combinatorial complexity when exposed to complex learning. It will not scale up to the real human brain. Elman (2005) is among the first to admit this. Still, SNR can be used to elucidate the general principles. Among such principles is abstract notion evolution from vague and fuzzy toward specific and concrete (Elman, 1993; Olguin & Tomasello, 1993; Tomasello & Olguin, 1993). In the following sections, we describe how dynamic logic systematically utilizes this principle. We also will address another important principle of mind organization brought up by Nolfi, Elman, and Parisi (1994): learning is motivated by internal drives. There is an important difference, however, between Elman's (2005) discussion of nonspecific emergence and the purposeful emergence mechanisms that we consider later: the instinct for knowledge.

Michael Tomasello (2001, 2003) suggests that the first principle of the human mind organization, the most important mechanism of the human brain required to learn language, is not language-specific but, more broadly, cultural and social. It is our ability to perceive other people as intentional agents. We understand that other people have intentions and plans to achieve them; we can figure out what these intentions and plans are. This is the foundation for our entire symbolic culture. The neural mechanisms of this ability are not known. How reasonable is it that we are born with an innate model for other people's intentions and plans? In the following sections, I describe a mathematical theory of joint learning of cognition and language. Its most important premise is that we are born with an innate drive, an instinct for knowledge. It determines the purposiveness of our existence, our higher mental abilities, and our ability to create symbolic culture. It is mathematically possible that a significant or even most important aspect of this drive is to acquire knowledge about other people's intentions and plans. It would be a fascinating enterprise to establish relationships between these two theories through laboratory psychological and neural research.

Let us summarize goals and achievements of cognitive linguistics. Connectionist architectures demonstrated learning of complicated syntax patterns without explicit rules and without explicit examples. They demonstrated elements of joint language learning and meaning creation (cognition). Still, these type architectures face CC and do not scale up. Motivational forces inherent to symbols, which were recognized by Saussure and analytic psychology, made inroads into linguistics and psychology. Still, symbols and signs continue to be mixed up.

The Mind: Concepts, Emotions, and Instincts

Difficulties of mathematical theories of the mind are not of purely mathematical origin. Their development began before the necessary intuitions of how the mind works became well-known to engineers. Newton, as often mentioned, did not consider himself as evaluating various hypotheses about the working of the material world; he felt that he possesses what we call today a physical intuition about the world (Westfall, 1981). An intuition about the mind points to mechanisms of concepts, emotions, instincts, imagination, behavior generation, consciousness and unconscious. Ideas of Dostoyevsky and Nietzsche and psychological theories of Freud and Jung, however, were too complex for many psychologists. It took a long time before they were considered part of science by engineers and mathematicians. An essential role of emotions in the working of the mind was analyzed from the psychological and neural perspective by Grossberg and Levine (1987), from the neurophysiological perspective by Antonio Damasio (1995), and from the learning and control perspective by the author (Dmitriev & Perlovsky, 1996; Perlovsky, 1998b, 1999). But the broad scientific community has been slow in adopting these results. One reason is the cultural bias against emotions as a part of thinking processes. Plato and Aristotle thought that emotions are bad for intelligence; this is a part of our cultural heritage ("one has to be cool to be smart"), and the founders of Artificial Intelligence repeated this truism about emotions even as late as the 1980s (Newell, 1983). Yet, as discussed in the next section, combining conceptual understanding with emotional evaluations is crucial for overcoming the combinatorial complexity as well as the related difficulties of logic.

Mechanisms of the mind, which seem essential to the development of the mathematical semantics, include instincts, concepts, emotions, and behavior. The mind serves for satisfaction of the basic instincts that have emerged as survival mechanisms even before the mind. What constitutes instincts and drives are topics of debates in psychological and linguistic communities (e.g., language instinct). For the purpose of developing a mathematical description (Perlovsky, 2004, 2006a), it is sufficient to consider instinct operations similar to internal sensors; for example, when a sugar level in blood goes below a certain level an instinct "tells us" to eat. To eat or to satisfy any bodily need, the mind has to understand the world around it. The need to understand drives cognition processes; I called it the knowledge instinct (Perlovsky, 2001; Perlovsky & McManus, 1991). It is described mathematically in the next section. A similar mechanism drives learning of language and can be called the language instinct. In this definition, the language instinct only indicates to our mind the basic need to understand and use language; it does not encompass specific mechanisms postulated by Chomsky (1972, 1981), Pinker (2000), Jackendoff (2002), or Tomasello (2001, 2003). These specific mechanisms still have to be elucidated, which constitutes a significant part of linguistic research (on both sides of the isle dividing or joining nativist and cognitivist approaches).

The most accessible to our consciousness is a mechanism of the mind, which operates with concepts. Concepts are like internal models of the objects and situations; this analogy is quite literal (e.g., during visual perception of an object, an internal concept-model projects an image onto the visual cortex, which is matched there to an image projected from the retina; this simplified description will be refined later). Mechanism of concepts evolved for instinct satisfaction; linking concepts and instincts in the mind involves emotions. Emotions are neural signals connecting instinctual and conceptual brain regions. Whereas, in colloquial usage, emotions often are understood as facial expressions, higher voice pitch, exaggerated gesticulation, these are the outward signs of emotions, serving communication. A more fundamental role of emotions within the mind system is that emotional signals evaluate concepts for the purpose of instinct satisfaction. This evaluation is not according to rules or concepts (like in rule-systems of artificial intelligence) but according to a different instinctual-emotional mechanism described next. The knowledge instinct and emotional mechanism are crucial for breaking out of the vicious cycle of combinatorial complexity; they lead to dynamic logic as discussed later.

Conceptual-emotional understanding of the world results in actions in the outside world or within the mind. We only touch on the behavior of improving understanding and knowledge of the language and the world. In the next section, we describe a mathematical theory of conceptual-emotional recognition and understanding. As we discuss, in addition to concepts and emotions, it involves mechanisms of intuition; imagination; and conscious, unconscious, and aesthetic emotion. This process is intimately connected to an ability of the mind to form symbols and interpret signs. The mind involves a hierarchy of multiple levels of concept-models, from simple perceptual elements (like edges or moving dots) to concept-models of objects, to complex scenes, and up the hierarchy toward the concept-models of the meaning of life and purpose of our existence. Parallel to this hierarchy of cognition, there is another hierarchy of language. Both hierarchies interact with each other and, although they have a degree of autonomy, cannot exist without this interaction. Psychological properties of symbols, which sometimes seem mysterious, emerge in this interaction of the two hierarchies. These interacting hierarchies are responsible for the tremendous complexity of the mind, yet relatively few basic principles of mind organization go a long way explaining this system.

Modeling Field Theory of Cognition

Modeling field theory (MFT) describes synaptic fields of neuronal activations, which implement model-concepts of the mind (Perlovsky, 2001). It is a multilevel, heterohierarchical system. The mind is not a strict hierarchy. There are multiple feedback connections among several adjacent levels; hence, the term *heterohierarchy*. MFT mathematically implements mechanisms of the mind in previous sections. At each level, there are concept-models generating top-down signals, interacting with input and bottom-up signals. These interactions are governed by the knowledge instinct, which drives concept-model learning, adaptation, and formation of new concept-models for better correspondence to the input signals.

This section describes a basic mechanism of interaction between two adjacent hierarchical levels of bottom-up and top-down signals (fields of neural activation); sometimes, it will be more convenient to talk about these two signal-levels as an input to and output from a (single) processing-level. At each level, output signals are concepts recognized in (or formed from) input signals. Input signals are associated with (or recognized or grouped into) concepts according to the models and the knowledge instinct at this level. The knowledge instinct is described mathematically as a maximization of similarity between the models and signals. In the process of learning and understanding input signals, models are adapted for better representation of the input signals so that similarity increases. This increase in similarity satisfies the knowledge instinct.

At each level, the output signals are concepts recognized (or formed) in input signals. Input signals **X** are associated with (or recognized or grouped into) concepts according to the representations-models and similarity measures at this level. In the process of association-recognition, models are adapted for better representation of the input signals, and similarity measures are adapted so that their fuzziness is matched to the model uncertainty. The initial uncertainty of models is high, and so is the fuzziness of the similarity measure; in the process of learning, models become more accurate and the similarity more crisp, and the value of the similarity measure increases. This mechanism is called *dynamic logic*.[2]

During the learning process, new associations of input signals are formed, resulting in evolution of new concepts. Input signals {**X**(n)} is a field of input neuronal synapse activation levels. Here and in the following, curve brackets {…} denote multiple signals: a field. Index n = 1,… N, enumerates the input neurons, and **X**(n) are the activation levels. Concept-models {$\mathbf{M}_h$(n)} are indexed by h = 1, … H. To simplify discussion, we talk later about visual recognition of objects; and we talk as if retina and visual cortex form a single processing layer (in fact, there are about 100 neuronal layers between retina and a visual cortex layer that recognizes objects). Each model $\mathbf{M}_h$(n) is a representation of the signals **X**(n) expected from a particular object, *h*. Each model depends on its parameters {$\mathbf{S}_h$}, $\mathbf{M}_h$($\mathbf{S}_h$,n). Parameters characterize object position, angles, lightings, and so forth. In this highly simplified description of a visual cortex, *n* enumerates the visual cortex neurons, **X**(n) are the bottom-up activation levels of these neurons coming from the retina through visual nerve, and $\mathbf{M}_h$(n) are the top-down activation levels (or priming) of the visual cortex neurons from previously learned object-models. Learning process attempts to match these top-down and bottom-up activations by selecting best models and their parameters.

Therefore, it is important to carefully define a mathematical measure of the best fit between models and signals; in other words, a similarity (or difference) measure between signals and models. In fact, any mathematical learning procedure, algorithm, or neural network maximizes some similarity measure or minimizes a difference. A difference measure used most often (for hundreds of years) is called least mean square; it is just an error between the model and signals (e.g., $\sum (\mathbf{X}(n) - \mathbf{M}_h(n))^2$); here, sum is taken over all signals *n*. This similarity measure, however, is only good for one model. When talking about the human mind, we need a similarity measure that can take into account multiple models in various combinations. We need a similarity measure between the sets of models and signals, *L* (or we can write explicitly the dependence of *L* on models and signals, L({**X**(n)},{$\mathbf{M}_h$(n)}). The similarity measure *L* also depends on model parameters and associations between the input synapses and concepts-models. It is constructed in such a way that any of a large number of object-models can be recognized. Correspondingly, a similarity measure is designed so that it treats each object model (or concept-model) as a potential alternative for each subset of signals

$$L(\{\mathbf{X}\},\{\mathbf{M}\}) = \prod_{n \in N} \sum_{h \in H} r(h)\, l(\mathbf{X}(n) \mid \mathbf{M}_h(n)); \qquad (1)$$

Let us explain this expression in a simple way. Here, $l(\mathbf{X}(n)|\mathbf{M}_h(n))$ (or simply $l(n|h)$) is called a conditional partial similarity, which means that it is just a similarity between one signal $\mathbf{X}(n)$ and one model $\mathbf{M}_h(n)$. Parameters r(h) are proportional to the number of objects described by the model *h*; they are not essential and used for convenience so that we can define each l(n|h) for a single object. Expression (1) accounts for all possible combinations of signals and models in the following way. Sum $\sum$ ensures that any of the object-models can be considered (by the mind) as a source of signal $\mathbf{X}(n)$. Product $\prod$ ensures that all signals have a chance to be considered. (Even if one signal is not considered, the entire product is zero, and similarity L is 0; so for good similarity all signals have to be accounted for. This does not assume exorbitant amount of attention to each minute detail; among models, there are vague simple models for everything else). In a simple case, when all objects are perfectly recognized and separated from each other, there is just one object-model corresponding to each signal (other l(n|h) = 0). In this simple case, expression (1) contains just one item, a product of all non-zero l(n|h). In general case, before objects are recognized, L contains a large number of all combinations of models and signals; a product over *N* signals is taken of the sums over *H* models, this results in a total of H^N items; this was the cause for the combinatorial complexity discussed previously.

Psychologically, maximization of similarity measure (1) is an instinct, an unconditional drive to improve the correspondence between input signals and internal representations-models. Let us emphasize once more that this instinct demands only maximizing one quantity: similarity *L*. The mathematical mechanisms of how this is achieved follows from the instinct structure (1) and are specified later. The knowledge about the world is contained in the models; therefore, similarity maximization is a mechanism of the knowledge instinct. Because models are adapted to input signals, knowledge depends on the realities of the surrounding world. Therefore, our knowledge is not a set of empty codes but represents objective reality. How good these representations are in individual minds is determined by a multitude of factors. In part, it is determined by the initial states of models. Some aspects of the models are inborn; others are acquired from culture, mostly from language, which we discuss later.

In the process of learning, concept-models constantly are modified. From time to time, a system forms a new concept while retaining an old one as well; alternatively, old concepts are sometimes merged. (Formation of new concepts and merging of old ones require some modification of the similarity measure (1), as discussed in Perlovsky [2001, 2006].)

The learning process consists of estimating model parameters $\mathbf{S}_h$ and associating subsets of signals with concepts by maximizing the similarity (1). Although (1) contains combinatorially many items, dynamic logic maximizes it without combinatorial complexity (Perlovsky, 1996b, 1997, 2001). First, fuzzy association variables f(h|n) are defined

$$f(h|n) = r(h)\, l(n|h) \Big/ \sum_{h' \in H} r(h')\, l(n|h'). \qquad (2)$$

These variables give a measure of correspondence between signal **X**(n) and model $\mathbf{M}_h$ relative to all other models, *h'*. A mechanism of concept formation and learning, the dynamics of modeling fields (MF) is defined by (2) along with the following equations,

$$\mathbf{S}_h = \mathbf{S}_h + \alpha \sum_n f(h|n)[\partial \ln l(n|h)/\partial \mathbf{M}_h] \partial \mathbf{M}_h / \partial \mathbf{S}_h, \tag{3}$$

$$r(h) = N_h / N; \; N_h = \sum_n f(h|n); \tag{4}$$

Here, parameter α determines the iteration step and speed of convergence of the MF system; N_h is as a number of signals **X**(n) associated with or coming from an object-model *h*. As already mentioned, in the MF dynamics, similarity measures are adapted so that their fuzziness is matched to the model uncertainty. Mathematically, this can be accomplished in several ways, depending on the specific parameterization of the conditional partial similarity measures, l(n|h); for example, they can be defined as familiar bell-shaped Gaussian functions,

$$l(n|h) = (2\pi)^{-d/2} (\det \mathbf{C}_h)^{-1/2} \exp\{- 0.5(\mathbf{X}(n) - \mathbf{M}_h(n))^T \mathbf{C}_h^{-1} (\mathbf{X}(n) - \mathbf{M}_h(n))\}. \tag{5}$$

Here, *d* is the dimensionality of the vectors **X** and **M**, and $\mathbf{C}_h$ is a covariance. These functions describe bell-shape forms centered at $\mathbf{M}_h$(n) with widths defined by $\mathbf{C}_h$. The dynamics of fuzziness of the MF similarity measures is defined as

$$\mathbf{C}_h = \sum_n f(h|n) (\mathbf{X}(n) - \mathbf{M}_h(n))(\mathbf{X}(n) - \mathbf{M}_h(n))^T / N_h. \tag{6}$$

Initially, models do not match data; covariances are large; bell-shapes are wide; and association variables, f(h|n), take homogeneous values across the data, associating all concept-models *h* with all input signals *n*. As matching improves, covariances become smaller; bell-shapes concentrate around the model-concepts $\mathbf{M}_h$(n); and the association variables, f(h|n), tend to high values 1 for correct signals and models and zero for others. Thus, certain concepts get associated with certain subsets of signals (objects are recognized and concepts formed). The following theorem was proven (Perlovsky, 2001).

> **Theorem:** Equations (2) through (6) define a convergent dynamic system MF with stationary states given by $\max_{\{\mathbf{S}_h\}} L$.

It follows that the previous equations indeed result in concept-models in the "mind" of the MFT system, which are most similar (in terms of similarity (1)) to the sensory data. Despite a combinatorially large number of items in (1), a computational complexity of the MFT is relatively low; it is linear in *N* and, therefore, could be implemented by a physical system like a computer or brain. (Let me emphasize that using Gaussian functions here is not like Gaussian assumption often used in statistics. Gaussian assumption assumes that the signals are Gaussian; this limits the validity of most statistical methods. Our similarity is quite general; it only assumes that the deviations between the models and signals are Gaussian; also, using many models in the sum in (1) can represent any statistical distribution.) Convergence of the MF system stated in the previous theorem assumes a definite set of input signals. In reality, new sensory signals reach our mind all the time; therefore, new concepts are continuously formed, and previously learned concepts are modified to fit new data.

To summarize, the knowledge instinct is defined by maximizing similarity (1). Its mathematical structure is chosen so that (prior to perception-cognition) any model-object can cause any signal. The mechanism of satisfying the knowledge instinct by maximizing (1), dynamic logic, is given by eqs. (2) through (4). Eqs. (5) and (6) are convenient, but their specific forms are not necessary.

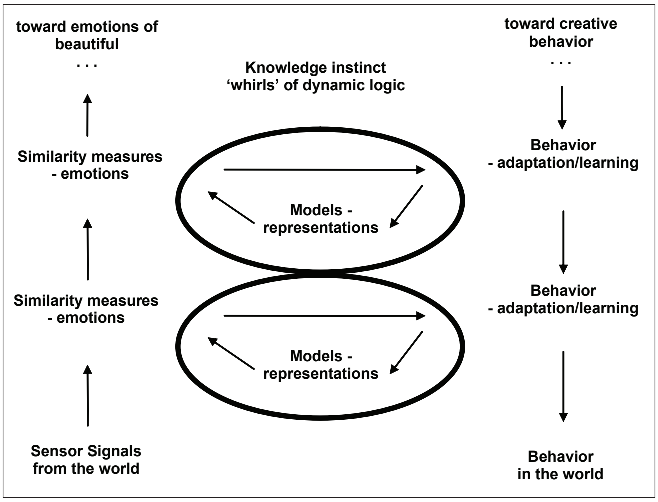
From a neurophysiological standpoint, neural signals relating concepts to instincts (similarities) are evaluative emotional signals. Emotional signals related to satisfaction of knowledge instinct mathematically are described by changes in similarity measure (1) during learning; eqs. (2) through (4). These emotions are directly related not to satisfying bodily needs but only spiritual needs for increased knowledge, and, according to Kant, they are called aesthetic emotions. Of course, there is no dualism; the knowledge instinct and aesthetic emotions are implemented in the brain neural mechanisms. These mechanisms, though, are removed from direct bodily needs, and in this sense, they can be called spiritual. Aesthetic emotions are not something specifically related to art; these emotions are involved in every act of perception and cognition.

Hierarchy of Cognition

The previous section described operation of a single MFT level, modeling a single level of the mind hierarchy. Like the mind, MFT is a heterohierarchical system consisting of multiple levels. Roughly speaking, at lower levels of the hierarchy are perceptual elements, objects; higher up are relationships among objects, situations, and more and more abstract and general model-concepts, and near the top are the most general concepts of the purpose and meaning of life. At every level, there is a similarity measure defining the knowledge instinct, models, emotions, and actions, including adaptation of models. An input to each level is a set of signals $\mathbf{X}(n)$. The result of signal processing at a given level are activated models, or concepts h recognized in the input signals n; these models, along with the corresponding instinctual signals and emotions, may activate behavioral models and generate behavior at this or lower levels. The activated models also send output signals from this level to the next processing level; these signals could be defined as model activation signals, a_h,

$$a_h = \sum_{n \in N} f(h|n). \quad (7)$$

Figure 1. Hierarchical MF system



Note: At each level of a hierarchy there are models, similarity measures, and actions (including adaptation, maximizing the knowledge instinct—similarity). High levels of partial similarity measures correspond to concepts recognized at a given level. Concept activations are output signals at this level, and they become input signals to the next level, propagating knowledge up the hierarchy. The hierarchy is not strict; interactions may involve several levels. At the top of the hierarchy, there are models of meaning and purpose, related emotions of beautiful, and creative behavior.

These signals, indicating recognized concept-models, become input signals for the next processing level, where more general concept-models are created. Hierarchical MF system is illustrated in Figure 1.

Operations of the knowledge instinct and dynamic logic are mostly unconscious. Only when concept-model is matched to a specific content and become crisp is it accessible to consciousness. More crisp and concrete models are more conscious (i.e., the mind, at will, can direct attention to, access, and operate with these models). Concept-models at lower hierarchy levels correspond to concrete objects. A child learns many of these models in the first months of life; they become concrete and conscious. Simple relations among objects, which are directly observable, also are learned early in life and become concrete and conscious. Learning of these models is said to be grounded in direct experience. Higher up in the hierarchy are more abstract cognitive models, which cannot be grounded in direct experience. Early in life, they remain fuzzy and less conscious. Still higher up, there are even more abstract models; some of them remain fuzzy and unconscious throughout life. Note that these models to attain crisp and conscious state have to be crisply and consciously related to the entire wealth of conceptual knowledge and experience at lower levels. People that are called knowledgeable and wise have more crisp and conscious models at higher levels of the mind hierarchy.

In the foundation of psyche, there are unconscious fuzzy models-archetypes. Every process of learning a concept-model involves a fuzzy unconscious model, which becomes more crisp and conscious and more clearly connected to experience and other concepts. This process connects conscious and unconscious and increases the limits of knowledge and consciousness; according to Carl Jung's (1921) definition, it is a symbol process. Here, I am using the notions of symbol and sign as used by Jung, Karl Pribram (1971), and general culture, and which is different from some definitions in classical semiotics and artificial intelligence. I'll continue discussion of motivations for various definitions in the last section. Here, I use the words *symbol* for adaptive processes creating new knowledge and *sign* for nonadaptive signals. The symbol process can take place completely inside the mind and does not have to involve signs in the outer world. Input signals from the lower level of the mind are signs on which the symbol process operates. Out of these signs, with the help of a fuzzy unconscious model, the symbol process creates a new concept at its hierarchical level, which is crisper and more conscious than the original fuzzy model. When the symbol-process ends, the result is a new sign, which can be used at a higher level in the hierarchy of the mind to create new symbols.

The higher in the mind of the hierarchy we attempt to extend this process, the less grounded in direct experience it becomes. The resulting concept-models could evolve in the following two ways. They remain fuzzier and less conscious than lower-level models grounded in direct experience, or they could become crisp and conscious agglomerations of arbitrary lower-level models that do not correspond to anything useful for human life and do not increase knowledge in any useful way. Increasing knowledge in a useful way is only possible due to language, as discussed in the following two sections.

Language MFT

Learning language, when described mathematically, using previously developed techniques, leads to combinatorial complexity. This is a general mathematical problem of learning algorithms that we discussed in sections 2 and 3. It is independent from specific mechanisms of language learning and equally applies to all theories of language learning. To overcome combinatorial complexity and to develop mathematical models adequate for language learning, we extend MFT developed for cognition in section 6. Like cognitive MFT described earlier, language is a hierarchical system; it involves sounds, phonemes, words, grammar, phrases, and sentences, and each level operates with its own models. Thus, we need to develop language models suitable for MFT and dynamic logic. In the human mind, these models to some extent are results of evolution; for computational intelligent systems, we have to develop them, and this development at each level is a research project, which is added by a number of already described language models in linguistics (Jackendoff, 2002; Mehler, 2002; Pinker, 2000; Rieger, 1981). A related challenge is to determine mechanisms of language evolution, so that specific mechanisms of contemporary languages evolve in processes of cultural evolution.

Here, I discuss an approach to the development of models of phrases from words. Given a large corpus of text, we would like to learn which word combinations are good models (i.e., used often and model most of the data). These models can be used for text understanding; for example, it could be used for an understanding-based search engine. There is a more general aspect of the development in this section; when combined with section 6, these techniques can be used to extend cognitive and language models to higher levels of a hierarchy and for integrating cognitive and language hierarchies addressed in sections 9 and 10. The difficulty of this task is related to the fact that phrases do not necessarily neatly follow one another, but they might overlap and form complex nested expressions. For example (Elman, 2003): "The man who came late is your uncle."

A simple way to learn this kind of sentences is to remember all kinds of sentences that are encountered in language. Clearly, humans can do better. The Chomsky (1972) approach was to figure out innate logical rules for syntax to deal with sentences like this. The Chomsky (1981) approach was to use parametric models (rules and parameters) instead of logical rules. Both approaches, as we discussed, faced combinatorial complexity. Elman demonstrated that a neural network can learn these types of sentences without ever encountering them. This neural network, however, was carefully constructed for a limited set of problems. It would not scale up to real language learning; it would face combinatorial complexity, as we already discussed.

Let us discuss MFT for learning phrases without CC. The input data, $\mathbf{X}(n)$ in this phrase-level MF system are word strings; for simplicity, of a fixed length, S, $\mathbf{X}(n) = \{ w_{n+1}, w_{n+2} \ldots w_{n+S} \}$. Here, w_n are words from a given dictionary of size K, $W = \{w_1, w_2 \ldots w_K\}$, and n is the word position in a body of texts. A simple phrase model is "a bag of word"; that is, a model is a subset of words from a dictionary without any order or rules of grammar,

$$\mathbf{M}^L_h(\mathbf{S}_h,n) = \{w_{h,1}, w_{h,2} \ldots w_{h,S}\}. \tag{7}$$

A superscript L here denotes a language model; the parameters of this model are its words, $\mathbf{M}^L_h(\mathbf{S}_h,n) = \mathbf{S}_h = \{w_{h,1}, w_{h,2} \ldots w_{h,S}\}$. The language learning (traditionally called language acquisition in Chomskyan linguistics) consists of defining models-concepts-phrases best characterizing the given body of texts in terms of a similarity measure.

Conditional partial similarities between a string of text, $\mathbf{X}(n)$ and a model $\mathbf{M}^L_h$ could be defined by a proportion of the matches between the two sets, $\mathbf{X}(n)$ and $\mathbf{M}^L_h$, $l(n|h) = |\mathbf{X}(n) \cap \mathbf{M}^L_h|/S$. Thus, similarity (1) is defined, and it could be maximized over the unknown parameters of the system, $\{ \mathbf{S}_h \}$; that is, over the word contents of phrases. Maximization of this language-similarity gives a mathematical formulation of the language instinct (Pinker, 2000). The language instinct mechanism is mathematically similar to the knowledge instinct; the main difference is that the language instinct maximizes similarity between language models and language data. Relations of language to cognition of the world is not considered within Chomskyan linguistics, and it is not a part of language instinct as formulated by Pinker (2000). We consider it in later sections.

Satisfaction of the language instinct, maximization of similarity between language-models ,and language signals result in language acquisition. Using the previously defined conditional similarities and phrase-models would result in learning models-phrases. The difficulty is that the dynamic logic, as described in the previous section, cannot be used for maximizing similarity. In particular, (3) requires evaluating derivatives, which requires a smooth dependence of models on their parameters. But bag-models do not depend smoothly on their word content. For example, a bag-model {Leonid, sit, chair} cannot be differentiated with respect to parameters *sit* or *chair*. This is a major difficulty: Any language model, at the level of phrases and above, is essentially a list, graph, or tree, which cannot be differentiated with respect to its word-content (or structure-content). Without dynamic logic, the computational complexity of similarity maximization becomes combinatorial ~ $K^{(H*N*S)}$; this is a prohibitively large number. This is the reason why old mathematical methods cannot be used for learning language and why computers do not talk and do not understand language yet.

The combinatorial complexity of this solution is related to a logic-type similarity measure, which treats every potential phrase-model (every combination of words) as a separate logical statement. The problem can be solved by extending dynamic logic as follows. Define the original vague state of phrase-models (phrase-potentialities) as long strings of words, much longer than actual phrases. During dynamic-logic processes, each vague model-phrase is compared to the corpus of text. At every iteration, a word least belonging to the phrase, on average over the text corpus, is determined and eliminated from the model. This procedure is qualitatively similar to differentiation, and it can be applied to discontinuous nondifferentiable functions, like sets of words (or other structural elements). Similar to section 6, original vague models poorly correspond to all phrases in the text corpus. As dynamic logic process progresses, models are becoming shorter and more specific; they become more selective, and they better correspond to specific phrases. This process ends with short specific model-phrases (phrase-actualities) corresponding to the content of the text corpus. This result is not unlike Elman (2003) with the difference that dynamic logic avoids CC can be scaled up and applied to the entire content of language.

We give now a mathematical description of this process. Define fuzzy conditional partial similarity measures (a similarity between one word sequence, $\mathbf{X}(n)$, and one model, $\mathbf{M}^L_h$):

$$l(n|h) = (2\pi\sigma_h^{\,2})^{-S/2} \exp\{- 0.5 \sum_s \; e(n,h,s)^2 / \sigma_h^{\,2} \}. \tag{8}$$

Here, e(n,h,s) is a distance (measured in the numbers of words) between the middle of the word sequence $\mathbf{X}(n)$ and the closest occurrence of the word $w_{h,s}$; the sum here is over words belonging to the phrase-model *h*. The search for the nearest word is limited to $\mathbf{X}(n)$ (S words), and e(n,h,s) falling outside this range can be substituted by a (S/2+1). Variance, determining fuzziness of this similarity measure, is given by a modification of (6),

$$\sigma_h^{\,2} = \sum_n \; f(h|n) \sum_s \; e(n,h,s)^2 / N_h. \tag{9}$$

Dynamic logic requires defining fuzzy contents of phrase-models, which can be done as follows. Define the average distance, δ, of the word $w_{h,s}$ from its phrase-model, *h*

$$\delta(h,s) = \sum_n \; f(h|n) \sum_s \; e(n,h,s)^2 / N_h; \tag{10}$$

This is an average distance over the entire text corpus. It is closely related to the measure of fuzzy phrase contents, or measure of belonging of the word *s* to phrase *h*. We define it as a probability-like measure of the word $w_{h,s}$ belonging to a model-phrase h, φ(s|h):

$$\varphi(s|h) = p(h|s) / \sum_{s' \in h} p(h|s'); \; p(h|s) = (2\pi\sigma_h^{\,2})^{-1/2} \exp\{-\,0.5 \sum_{s} \delta(h,s) / \sigma_h^{\,2}\}, \tag{11}$$

The last equation here is a bell-shaped curve, a nonnormalized measure of belonging of word *h* to phrase *s*; the first equation gives φ, a probability-like normalized measure for the word *s* relative to all other words in the model *h*. This measure is used now to define the dynamics of the word contents of the phrase-models in the dynamic-logic process as follows. Let us limit the problem to learning phrases of a fixed length, say, we would like to learn five-word phrases. Start with a large value of S >> 5 (e.g. S = 50) and with arbitrary word-contents of phrases (7). On each iteration, compute eqs. (8) through (11). Reduce S by 1; in each phrase-model eliminate one word with the minimal φ(s|h),

$$\text{for each } h, \text{ find } s' = \operatorname{argmin}_s \varphi(s|h), \tag{12}$$

$w_{h,s'}$ is the least probable word in model *h*, and it is eliminated on this iteration. *S* is changed to *S-1*. Continue iterations until *S* reaches the desired value 5.

The dynamics defined in this way result in learning phrase-models and accomplishes the goal of language acquisition without combinatorial complexity. The computational complexity is moderate, ~ N*H*K*S. This overcoming of CC is the major goal of this section. Limitations of the previous procedure, like predefined length of phrases, can be overcome similar to the discussion in the section Modeling Field Theory of Cognition (see also Perlovsky, 2001).

The bag-of-word phrase models considered previously are simpler than known structures of natural languages with treelike dependencies, syntactic rules, and word order (Jackendoff, 2002; Mehler, 2002; Pinker, 2000; Rieger, 1998). These more complicated real linguistic models can be used in place of a simple distance measure e(n,h,s) in (8). This does not lead to a significant growth of complexity. In this way, the models of noun and verb phrases and tree structures can be incorporated into the previous formalism. One of the challenges of contemporary linguistics is to identify which aspects of the models are innate so that every child learns a human language, and to identify which aspects are learned so that any of thousands of languages can be learned. It is quite possible that the inborn, innate information about a conceptual structure of language is contained in simple bag-type models of the type considered previously; the rest could be learned jointly with cognition, as considered later. We do not consider here emotional content of language (Perlovsky, 2006b).

The procedure outlined in this section is general in that it is applicable to all higher levels in the mind hierarchy. Lower-level models may require continuous parametric models, like laryngeal models of phonemes (Lieberman, 2000). These can be learned from language sounds using procedures similar to the section Modeling Field Theory of Cognition. Higher hierarchical models, like models of phrases, or language models corresponding to complex abstract concepts contained in paragraphs, or books, are learned from lower-level models using the technique described in this section. This is also true about high-level cognitive models of relationships among objects, situations, and so forth. Are we born with complex innate mechanisms of these models (using structured sets or graphs), or are simple bag-models sufficient? This is a challenge for linguistics today. The general approach described in this section overcomes CC of learning algorithms and can be used with a variety of specific language learning models.

Integrating Language and Cognition

Let me repeat that today, we still do not know neural mechanisms combining language with thinking or their locations in the brain. Mathematical mechanisms discussed for unifying cognition and linguistics (Brighton et al., 2005; Christiansen & Kirby, 2003; Elman, 1996; Jackendoff, 2002) face combinatorial complexity for the same mathematical reasons that cognitive and language algorithms did in the past. Here we extend MFT to unifying cognition and language, while avoiding CC. We discuss a relatively simple mechanism that might be sufficient for joint learning of language and cognition and which corresponds to existing knowledge and intuition about these processes.

Integration of language and cognition in MFT is attained by integrating cognitive and language models (Perlovsky, 2002m 2004) so that a concept-model $\mathbf{M}_h$ is given by

$$\mathbf{M}_h = \{ \mathbf{M}^C_h , \mathbf{M}^L_h \}; \tag{12}$$

Here, $\mathbf{M}^C_h$ denotes a cognitive part of the model of an object or situation in the world, and $\mathbf{M}^L_h$ is a linguistic part of the model. Consider now this integrated model as the mind's mechanism of integrating language and cognition. A data stream constantly comes to mind from all sensory perceptions; every part of this data stream is evaluated constantly and associated with models (12) according to the mechanisms of dynamic logic described in previous sections. In this fuzzy dynamic association, at the beginning, the models are fuzzy; cognitive models vaguely correspond to uncertain undifferentiated sensory perceptions. Language models vaguely correspond to sounds. This is approximately a state of mind of a newborn baby. First, models of simple perceptions differentiate; objects are distinguished in visual perception. Language sounds are differentiated from other sounds. In (12), some cognitive models become crisper than other cognitive models. Until about one year of age, perception models corresponding to simple objects become crisper at a faster rate than language models.

Gradually, models are adapted, their correspondence to specific signals improve, selectivity to language signals and nonlanguage sounds are enhanced. Language models are associated with some degree of specificity with words (sentences, etc.), and cognitive models are associated with objects and situations of perception and cognition. Between the first and second year of life, the speed of adaptation of language models tremendously accelerates and overtakes learning of cognitive models.

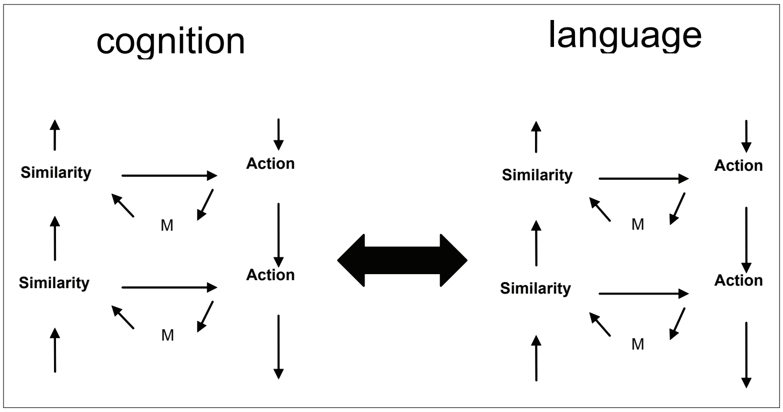
Some degree of association between language and cognitive models occurs before any of the models attain a high degree of specificity that is characteristic of the grown-up conscious concepts. Language and cognition are integrated at a preconscious level. Certain language models evolve faster than their corresponding cognitive models and vice versa. Correspondingly, uncertainty and fuzziness of the two aspects of integrated models may differ significantly. Still, existence of a low-fuzzy linguistic model speeds up learning and adaptation of the corresponding cognitive model and vice versa. I suggest that this is a mechanism of interaction between language and cognition. Both abilities enhance each other.

The described mechanism of interaction between language and thinking may apply to ontological development and learning, biological specie evolution, and evolution of cultures. The differences between these learning and evolution processes are in the degree of specificity of a priori models (inborn, or accumulated in culture) and in the type of data available for learning and evolution. For example, child learning occurs in parallel in three realms: (1) linguistic models are learned to some extent independently from cognition, when linguistic data are encountered for the first time with limited or no association with perception and cognition (like in a newborn baby); (2) similarly, cognitive models can be learned to some extent independently from language, when perception signal data are encountered for the first time in limited or no association with linguistic data; and (3) linguistic and cognitive models are learned jointly when linguistic data are present in some association with perception signals; like during mother talking to a baby: "this is a car" (visual-perception-models and the corresponding linguistic-word-models are engaged together); another example is more complicated conversations: "Look at Peter and Ann, they are in love" (leads to learning related cognitive-models and phrase-models). The most significant part of learning, it seems, involves independent learning of language and cognitive parts of models when situations and their language descriptions are encountered independently from each other. Structure (12) provides for a cognitive placeholder fuzzy model for each language model, and vice versa. In this way, both types of models are learned gradually, always remaining associated; cognition helps language, and language helps cognition. In this way, knowledge is accumulated through generations.

MFT Hierarchical Organization

The previous section described a single processing level in MFT system integrating language and cognition. This mechanism of integrated models can integrate cognitive and language hierarchies, as illustrated in Figure 2. An amazing aspect of the human mind is that these two hierarchies are integrated in such a way that relationships among constituent models are preserved. For example, a cognitive model of a situation and the corresponding phrase model are constituted from lower-level models: objects and words. Correspondence between these objects and words in the object-word level is the same as between them, when they become constituent parts of the phrase-situation level model. And this holds true across a tremendous number of the phrase-situation level models, using various combinations of the same words from the lower level. This amazing property of our mind seems so obvious that nontrivial complexity of the required mechanism was noticed only recently (Deacon, 1998). The only mathematical description of a mechanism that promises such integration of the two hierarchies without CC is given in sections following (Perlovsky, 2002, 2004).

Figure 2. Hierarchical integrated language-cognition MF system



Note: At each level in a hierarchy there are integrated language and cognition models. Similarities are integrated as products of language and cognition similarities. Initial models are fuzzy placeholders, so integration of language and cognition is subconscious. Association variables depend on both language and cognitive models and signals. Therefore, language model learning helps cognitive model learning and vice versa. Abstract cognitive concepts are grounded in abstract language concepts.

Deacon (1998) suggested that the ability for two hierarchies sets the human mind apart from the rest of the animal world. For example, a dog can learn to bring shoes to a human master on a verbal command. A dog, it seems, can jointly learn language and cognition (a word *shoes* and an object *shoes*). This is only true, however, at the lower levels of the mind hierarchy, at the level of objects. The dog can do it because it perceives objects (shoes) in the world. Learning of a word-concept, shoes, is grounded in direct perception of objects in the world. Note that such a direct grounding in sensory signals exists only at the very bottom of the mind hierarchy. At higher levels, cognitive concepts are grounded in lower-level concepts. These higher levels exist only in the human mind. Try to teach a dog to understand the word *rational* or any abstract concept, which meaning is based on several hierarchical levels; this is not possible. It is known that the smartest chimps after long training barely can understand few concepts at the second level (Savage-Rumbaugh & Lewine, 1994).

Ability for learning higher levels of the hierarchy, it seems, is closely related to ability for language. The reason is that otherwise, learning of cognitive models does not have a ground for learning; there are no abstract concepts that could be directly perceived in the world. The only ground for learning abstract cognitive concepts is language concepts, which are learned from surrounding language and culture at many hierarchical levels. In an integrated MFT system, abstract cognitive models at higher levels in the hierarchy are grounded in abstract language models. Due to integration of language and cognition, language provides grounding for abstract higher cognitive models.

Cognitive models that proved useful in life and evolution cannot be transferred directly to the minds of the next generation. Cognitive models created by each generation are accumulated in culture and are transferred to the next generation through language. Cultural evolution gradually selects useful models. Language accumulates cultural knowledge at all levels in hierarchy of the mind.

Mechanisms of integration of cognition and language given by dual models, eq. (12), and dual hierarchies, Figure 2, are as if a bridge exists between nativist and cognitive linguistic approaches. Mathematical mechanisms proposed here can be used in conjunction with other proposed mechanisms, language specific or not. Many of the mechanisms for language and cognition discussed in literature (Chomsky, 1995; Elman et al., 1996; Jackendoff, 2002; Lieberman, 2000; Pinker, 2000; Tomasello, 2003) can be integrated with MFT structure discussed previously and take advantage of the dynamic logic overcoming CC.

Cognitivist approach rejects specific language mechanisms; for example, Tomasello (2003) suggests that understanding other people's intentions and goals is sufficient to acquire language and to create culture. Nativist approach seeks to explain language independently from cognition (Pinker, 2000). Consider the fact that some people master language very well, while other people are inept. Opposite examples also abound. This consideration seems to support some separation between language and cognition of people's intents. It is quite possible that cognitive mechanisms for inferring other people's intents proposed by Tomasello can incorporate these differences in speed of learning of these two abilities. Following Hurford (2001), I would like to mention that the strong polemical tone in some of the linguistic literature is symptomatic of a schism in modern linguistics, which hopefully can be resolved soon. Controversies between algorithmic and nonalgorithmic, learned vs. innate and instinctual, it seems, often are based on old divisions and definitions of terms rather than on actual differences among contemporary researchers about importance of various mechanisms. When laboratory studies will be combined with mathematical models capable of scaling up to the real mind; and when the model predictions will be tested against the experimental data, current divisions will yield to more interdisciplinary studies and intergroup cooperation.

Cultural evolution of language and cognition as well as ontological learning by a child could be supported by similar mechanisms. It is quite possible that a significant part of conceptual cognitive and language abilities (from words and objects up the mind hierarchy toward complex abstract concepts) can evolve and be learned based on few inborn mechanisms described in this chapter: MFT structure, the knowledge instinct, dynamic logic, dual model (12), and hierarchy (Figure 2). For example, Brighton et al. (2005) demonstrated that combinatorial compositionality of language emerges under proper conditions from a single simple mechanism. The main requirement is a mind's ability to guess-predict sounds for new situations from previous experiences (so that new sounds are understandable by the rest of the community). This property of accurate guessing is inherent to the MFT mechanisms, because dynamic logic evolves MFT structure from vague and fuzzy toward probabilistic and maximum likelihood. (The maximum likelihood principle is mathematically equivalent to the minimum description length principle used by Brighton et al., 2005). Implementing Brighton et al.'s approach with MFT will overcome current CC of that work. Also, like much of contemporary work on language evolution, Brighton et al. assumed preexisting meanings (i.e., cognition). Current effort is directed at overcoming these limitations toward joint evolution of language and cognition using MFT and dynamic logic (Perlovsky & Fontanari, 2006). It would be interesting to further connect language evolution to Elman's (2003) work on learning of complex syntax with relatively simple innate models. It seems that Elman's models can be mapped to the MFT architecture in a relatively straightforward way. I would emphasize that postulating one assumption (like innateness vs. learning) to explain that one fact does not lead too far. The essence of scientific theory emphasized by many linguistic researchers is in explaining many facts with few assumptions. The next section makes a step in this direction using the theory of dual language-cognition models and hierarchy to explain complex interrelations among language, cognition, and symbols.

Language, Cognition, and Symbols

Why is the word *symbol* used in such a different way—to denote trivial objects like traffic signs or mathematical notations and also to denote objects affecting entire cultures over millennia, like Magen David, Swastika, Cross, or Crescent?

Let us compare in this regard opinions of two founders of contemporary semiotics: Charles Peirce (19^{th} to 20^{th} century) and Ferdinand De Saussure (1916). Peirce classified signs into symbols, indexes, and icons. Icons have meanings due to resemblance to the signified (objects, situations, etc.); indexes have meanings by direct connection to the signified, and symbols have meaning due to arbitrary conventional agreements. Saussure used different terminology; he emphasized that the sign receives meaning due to arbitrary conventions. Saussure chose the term *sign* over *symbol* because the latter implies motivation. It was important for him that motivation contradicted arbitrariness.

Choice of convention for the most fundamental terms in semiotics is not arbitrary but ought to be motivated by understanding of the working of the mind and by the most widely used conventions across the culture. For this purpose, it is not irrelevant to note that Peirce considered himself a logician (logic implies arbitrariness of conventions), and in his personal life he abstracted himself from cultural conventions. Saussure was a linguist, he was better attuned to cultural conventions, and he was more sensitive to the fact that the word *symbol* implied nonarbitrary motivations.

Both Peirce and Saussure wanted to understand the process in which signs acquire meanings. Both of them failed; workings of the mind were not known at the time. Consider Peircian icons; they resemble objects or situations because of specific mechanisms of perception and recognition in our minds. These mechanisms should be analyzed and understood as an essential part of meaning creation. Peircian assumption that icons in themselves resemble situations in the world is too simplistic. Algorithms based on this assumption led to irresolvable difficulties related to combinatorial complexity. Similarly, arbitrariness emphasized by Peirce and Saussure did not help in understanding algorithms of meaning creation. Since arbitrary conventions also are expressed through signs, all signs get their meanings only in relation to or in contrast with other signs in a system of signs. Arbitrary signs, therefore, have no grounding in the real world. Meanings cannot be created by unmotivated choices on the interconnections of arbitrary signs; this type of choice leads to combinatorial complexity. In infinite systems, they lead to Gödelian contradictions. Similarly, mechanisms of meaning creation were not found by founders of symbolic AI when they used motivationally loaded word *symbol* for arbitrary mathematical notations. Mathematical notations, just because they are called symbols, do not hold a key to the mystery of cultural and psychological symbols.

Multiple meanings of the word *symbol* misguided their intuition. This is an example of what Wittgenstein called "bewitchment by language."

The MF theory and dynamic logic emphasize that meanings are created in processes connecting conscious and unconscious. There are two fundamental processes meaning creation in evolution of language and culture: differentiation and synthesis. First, differentiation consists of bringing unconscious into consciousness. It acts at the deepest levels of bringing unconscious archetypes into consciousness as well as at everyday levels of differentiating multiple aspects of various concepts and making these aspects more concrete and more conscious. This process takes millennia, and its results are stored in language. Its mathematical mechanisms are described in sections 6 and 7. Second, synthesis consists of connecting differentiated conscious concepts in language with cognition and through cognition with unconscious instinctual needs. Its mathematical mechanisms are described in sections 9 and 10.

Whereas differentiation is the essence of cultural and cognitive development, synthesis creates necessary conditions for differentiation. Both processes are necessary, yet their relationships are not simple. There is synergism but also opposition between differentiation and synthesis. The reason is that too strong a synthesis stifles differentiation: If the language hierarchy is not sufficiently vague, if it is too crisp, language may strongly predetermine meanings of cognitive concepts so that creation of new meanings is difficult and culture stagnates. The opposite side of the story is that differentiation can overtake synthesis. A large number of finely differentiated concepts might be created in language, but individual minds lag in their capacity for synthesis, for connecting language to cognition, and to essential demands of life. If this condition predominates in the entire culture, its meaning is lost for the people, and culture disintegrates. This was the mechanism of death of many ancient civilizations. Currently, predominance of synthesis is characteristic of certain Eastern cultures, whereas predominance of differentiation is characteristic of Western cultures. This direction for future research requires going beyond conceptual contents of languages and to study their emotional, motivational contents (Perlovsky, 2006b).

Both differentiation and synthesis are motivated by the instinct for knowledge. The motivated meaning creation, connecting conscious and unconscious, is consistent with Jungian explanations of the nature of symbols (1921). This motivates me to use the word *symbol* for the processes of meaning creation and to use the word *sign* for conventional or nonadaptive entities. I would also add, as a motivation for other semioticians (pun intended) to adopt these conventions, to entertain the following question: Why does the word *semiotics* leave a bitter taste in the minds of many physicists, engineers, and analytical psychologists, despite the obvious importance of the field? This could not be understood from the point of view of arbitrariness of conventions. Researchers, whose subjects are connected to the real world outside and inside human psyche might be repulsed by arbitrariness of the most important definitions. Founders of symbolic artificial intelligence were captivated by mathematical logic; they were not attuned to the fact that mathematical notations called symbols are not at all similar to psychological symbol-processes in the mind. I think this is the reason why, despite Gödel's results, proving inconsistency of logic, they still used formal logic to model the mind.

Let me summarize. In the context of the discussions in this chapter, a sign means something that can be interpreted to mean something else (like a mathematical notation, a word, or a traffic sign). The process of sign interpretation is a symbol-process, or symbol. This process resides in our minds. Interpretation or understanding of a sign by the mind according to MFT is due to the fact that a sign (e.g., a word) is a part of a model. The mechanism of sign interpretation is motivated by the knowledge instinct, which activates the model and connects the sign to the world outside and inside us. Second, a sign is understood in the context of a more general situation in higher levels of the mind hierarchy, containing more general concept-models. Recognized signs, which are the results of symbol processes, comprise input signals for the next level models, which cognize more general concept-models. Signs, therefore, are not just objects in the world but also are neural signals in our minds to which meanings are fixed as a result of the concluded symbol-processes. This corresponds to Pribram's (1971) interpretation of signs as nonadaptive neural signals with fixed meanings.

Meanings are created by symbol-processes in the mind. Language plays a special role in these processes. Language accumulates cultural knowledge of the world. Through communication among people, language provides grounding for abstract model-concepts at higher levels in the mind hierarchy. The mechanism of this relationship between language and cognition is joint language-cognitive models. These joint models are organized in parallel hierarchies of language models (words, texts) and cognitive models (world representations in the mind). Near the bottom of these hierarchies, words refer to objects. Higher up, complex texts refer to complex situations. An amazing result of the described mechanism is that words within texts refer to objects within situations, and this reference at higher levels corresponds to the words-objects relationships at lower levels. Because of this multilevel hierarchical structure maintaining meaningful relationships throughout the hierarchy, language is a coherent structure and not a set of arbitrary notations for arbitrary relationships. This meaning-maintaining hierarchy makes possible "the infinite use of finite means."

Cultural evolution results in selection and preservation in language of important meanings. They are related to concept-models important for cognition (and survival). Of course, at every given moment in cultural evolution, there are important and less important models. There are no simple measures for importance of meanings of various models and texts. But the deconstruction idea that meanings are arbitrary is unscientific. Scientific quest is to explain creation of meanings, and this chapter made a step in this direction.

In the early 1800s, Wilhelm von Humboldt (1999) suggested that languages, in addition to their explicit conventional outer form, also contain inner form full of potential and creativity. The mechanism of dynamic logic explains that the creative aspect of language exists in the integrated relationship between language and thinking; concept-models and meanings are developed unconsciously in interaction with language models. This process involves the knowledge and language instincts and aesthetic emotions related to satisfaction of these instincts.

A symbol-process involves conscious and unconscious; concepts and emotions; inborn models-archetypes; and models learned from culture, language, and cognition. Symbol processes continue up and up the hierarchy of models and mind toward the most general models. Due to language, they persist in culture through many generations. In semiotics, this process is called *semiosis*, a continuous process of creating and interpreting the world outside (and inside our minds). Symbols are processes creating meanings.

References

Bellman, R. E. (1961). *Adaptive control processes*. Princeton, NJ: Princeton University Press.

Bonnisone, P. P., Henrion, M., Kanal, L. N., & Lemmer, J. F. (1991). *Uncertainty in artificial intelligence 6*. North Holland, Amsterdam.

Brighton et al. (2005). Language as an evolutionary system. *Phys. Life Rev., 2*(3), 185-222.

Carnap, R. (1928/2003). *The logical syntax of language*. (trans. A. Smeaton). Open Court Publishing Company

Chomsky, N. (1972). *Language and mind*. New York: Harcourt Brace Javanovich.

Chomsky, N. (1981). Principles and parameters in syntactic theory. In N. Hornstein, & D. Lightfoot (Eds.), *Explanation in linguistics: The logical problem of language acquisition*. London: Longman.

Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT Press.

Christiansen, M. H., & Kirby, S. (2003). *Language evolution*. New York: Oxford University Press.

Damasio, A. R. (1995). *Descartes' error: Emotion, reason, and the human brain*. New York: Avon.

Deacon, T. W. (1998). *The symbolic species: The co-evolution of language and the brain*. New York: Norton.

Derrida, J. (1967). *Of grammatology*. Baltimore, MD: John Hopkins University Press.

Derrida, J. (1978). *Writing and difference*. Chicago: Chicago University Press.

De Saussure, F. (1916/1965). *Course in General Linguistics*. New York: McGraw-Hill.

Dmitriev, V. A., & Perlovsky, L. I. (1996). Art form as an object of cognitive modeling (towards development of Vygotsky's semiotics model). *Proceedings of the Conference on Intelligent Systems and Semiotics '96,* Gaithersburg, Maryland (pp. 385-389).

Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14*, 179-211.

Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition, 48*, 71-99.

Elman, J. L. (2003). Generalization from sparse input. *Proceedings of the 38th Annual Meeting of the Chicago Linguistic Society,* Chicago, Illinois.

Elman, J. L. et al. (1996). *Rethinking innateness: A connectionist perspective on development (neural networks and connectionist modeling)*. Cambridge, MA: MIT Press.

Fiske, J. (1987). Film, TV and the popular. *Continuum: The Australian Journal of Media & Culture, 12*, 4-5.

Gödel, K. (1994). *Collected works: Publications 1929-1936* (Vol. I). New York: Oxford University Press. (See also discussions in Penrose, R. (1994). *Shadows of the mind*. Oxford: Oxford University Press.)

Grossberg, S. (1982). *Studies of mind and brain*. Dordrecht, Holland: D. Reidel Publishing Co.

Grossberg, S., & Levine, D.S. (1987). Neural dynamics of attentionally modulated Pavlovian conditioning: Blocking, inter-stimulus interval, and secondary reinforcement. *Psychobiology*, *15*(3), 195-240.

Hilbert, D. (1928/1967). The foundations of mathematics. In J. van Heijenoort (Ed.), *From Frege to Gödel* (p. 475). Cambridge, MA: Harvard University Press.

Hurford, J. R. (2001). Review of P. Lieberman, 2000: Human language and our reptilian brain. *Quarterly Review of Biology, 76*(3), 383.

Hurford, J. R., Studdert-Kennedy, M., & Knight, C. (Eds.) (1998). *Approaches to the evolution of language: Social and cognitive bases*. Cambridge: Cambridge University Press.

Jackendoff, R. (1983). *Semantics and cognition*. Cambridge, MA: MIT Press.

Jackendoff, R. (2002). *Foundations of language: brain, meaning, grammar, evolution*. New York: Oxford University Press.

Jung, C. G. (1921). *Psychological types* (Vol. 6, Bollingen Series XX, 1971). Princeton, NJ: Princeton University Press.

Kecman, V. (2001). *Learning and soft computing: Support vector machines, neural networks, and fuzzy logic models (complex adaptive systems)*. Cambridge, MA: MIT Press.

Lakoff, G. (1987). *Women, fire, and dangerous things. What categories reveal about the mind*. Chicago: University of Chicago Press.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Langacker, R. W. (1987). *Foundations of cognitive grammar, Vol. 1: Theoretical prerequisites*. Stanford: Stanford University Press.

Lieberman, P. (2000). *Human language and our reptilian brain: The subcortical bases of speech, syntax, and thought*. Cambridge, MA: Harvard University Press.

Martinovsky, B. (1995). *Three cognitive theories.* Retrieved from http://www.ling.gu.se/~biljana/cv2002.html

Minsky, M. (1988). *The society of mind*. New York: Simon and Shuster.

Minsky, M. L. (1968). *Semantic information processing*. Cambridge, MA: MIT Press.

Nevatia R., & Binford, T.O. (1977). Description and recognition of curved objects. *Artificial Intelligence*, *8*(1), 77–98.

Newell, A. (1983). Intellectual issues in the history of artificial intelligence. In F. Machlup, & U. Mansfield (Eds.), *The study of information*. New York: J. Wiley.

Nietzsche, F. (1876). *Untimely meditations, sect. 5, Richard Wagner in Bayreuth* (Trans. Hollingdale). Cambridge: Cambridge University Press.

Nolfi, S., Elman J. L., & Parisi, D. (1994). Learning and evolution in neural networks. *Adaptive Behavior, 3*(1), 5-28.

Olguin, R., & Tomasello, M. (1993). Twenty-five-month-old children do not have a grammatical category of verb. *Cognitive Development, 8*, 245-272.

Peirce, C. S. (1897, 1903). Logic as semiotic: The theory of signs. In Buchler (Ed.), *The philosophical writing of Peirce* (1955), Dover, NY.

Peirce, C. S. (1935–1966). *Collected papers of Charles Sanders Peirce*. Cambridge, MA: Harvard University Press.

Perlovsky, L. I. (1987). *Multiple sensor fusion and neural networks. DARPA neural network study*. Lexington, MA: MIT/Lincoln Laboratory.

Perlovsky, L. I. (1988). Neural networks for sensor fusion and adaptive classification. *Proceedings of the 1st Annual International Neural Network Society Meeting,* Boston, Massachusetts (p. 42).

Perlovsky, L. I. (1996a). Gödel theorem and semiotics. *Proceedings of the Conference on Intelligent Systems and Semiotics '96,* Gaithersburg, Maryland (pp. 14-18).

Perlovsky, L. I. (1996b). Mathematical concepts of intellect. *Proceedings of the World Congress on Neural Networks,* San Diego, California (pp. 1013-1016).

Perlovsky, L. I. (1997). Physical concepts of intellect. *Russian Academy of Sciences, 354*(3), 320-323.

Perlovsky, L. I. (1998a). Conundrum of combinatorial complexity. *IEEE Trans. PAMI, 20*(6), 666-670.

Perlovsky, L. I. (1998b). Cyberaesthetics: Aesthetics, learning, and control. *Proceedings of the STIS'98,* Gaithersburg, Maryland.

Perlovsky, L. I. (1999). Emotions, learning, and control. *Proceedings of the International Symposium on Intelligent Control, Intelligent Systems & Semiotics,* Cambridge, Massachusetts (pp. 131-137).

Perlovsky, L. I. (2001). *Neural networks and intellect*. New York: Oxford University Press.

Perlovsky, L. I. (2002). Computational Linguistics, Semiotics and the Mind. *Proceedings of the Computational Linguistics Colloquium,* Trier, Germany.

Perlovsky, L. I. (2004). Integrating language and cognition. *IEEE Connections*, *2*(2), 8-12.

Perlovsky, L. I. (2006a). *Modeling field theory for higher cognitive functions*. Hershey, PA: Idea Group.

Perlovsky, L. I. (2006b). *The knowledge instinct*. New York: Basic Books.

Perlovsky, L. I., & Fontanari, J. F. (2006). How language can guide intelligence. *Proceedings of the EVOLANG'06,* Rome, Italy.

Perlovsky, L. I., & McManus, M. M. (1991). Maximum likelihood neural networks for sensor fusion and adaptive classification. *Neural Networks, 4*(1), 89-102.

Perlovsky, L. I., & McManus, M. M. (1991). Maximum likelihood neural networks for sensor fusion and adaptive classification. *Neural Networks, 4*(1), 89-102.

Pinker, S. (2000). *The language instinct: How the mind creates language*. Harper Perennial.

Pribram, K. (1971). *Languages of the brain*. New York: Prentice Hall.

Rieger, B. B. (Ed.). (1981). *Empirical semantics II. A collection of new approaches in the field* (vol. 13). Bochum, Germany: Brockmeyer.

Rieger, B. B. (1998). Tree-like dispositional dependency structures for non-propositional semantic inferencing: A SCIP approach to natural language understanding by machine. *Proceedings of the 7th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU-198),* Paris (pp. 351-358).

Rieger, B. (2002). Private communication.

Russell, B. (1919). *Introduction to mathematical philosophy*. London: George Allen and Unwin.

Savage-Rumbaugh, E. S., & Lewine, R. (1994). *Kanzi: The ape at the brink of the human mind*. New York: J. Wiley.

Talmy, L. (1988). The relation of grammar to cognition. In Rudzka-Ostyn (Ed.), *Topics in cognitive linguistics* (pp. 165–205). Amsterdam: Benjamins.

Tomasello, M. (2001). *The cultural origins of human cognition*. Cambridge, MA: Harvard University Press.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Tomasello, M., & Olguin, R. (1993). Twenty-three-month-old children have a grammatical category of noun. *Cognitive Development, 8*, 451-464.

Van Gelder, T., & Port, R. (1995). It's about time: An overview of the dynamical approach to cognition. In R. S. Westfall (Ed.), *Mind as motion: Explorations in the dynamics of cognition*. Cambridge, MA: MIT Press.

von Humboldt, W. (1999). Humboldt: On language. In M. Losonsky (Ed.), *On the diversity of human language construction and its influence on the mental development of the human species*. Cambridge: Cambridge University Press.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control, 8*, 338-352.

Endnotes

1 R. Jackendoff works within both paradigms, nativist, and cognitive toward unifying both methods.

2 Dynamic logic should not be confused with dynamic system theory used by some authors to describe cognition (see Van Gelder & Port, 1995). Dynamic systems theory usually describes a single process that occurs with limited (or no) interactions with other processes. When mathematics of dynamic systems is used to describe multiple interacting processes in conjunction with adaptation or learning, it leads to combinatorial complexity. Dynamic logic was specifically designed to address multiple concurrent interacting processes.