

**ON THE WAY TO SOLVING THE PROBLEM  
OF “SEMANTIC WEB – DATA BASE”**

**Y.I. KISLENKO**

**Abstract.** The present state of the problem of “Semantic Web – Data Base” is analyzed. “Semantic Web” is analyzed from the standpoint of the integration approach, covering the results of research in the fields of neurophysiology, psychology, philosophy; it allows to formally define the quantum of knowledge as a separate situation of the visual level and clearly define the scheme of its verbalization in the form of a basic semantic-syntactic structure. The main result of the integration approach is presented by the following thesis: “The structural level of the linguistic organization is derived from the structural and functional level of the neural organization of the visual path”. From here we have a productive conclusion: the structural and functional level of the language organization will be the same for all languages. The second component of the “Data Base” covers (should cover) the entire social cognitive potential of knowledge, presented by a plurality of accumulated texts. The structured level of knowledge base organization is presented by a very small fragment of the neural network, which reproduces a separate situation of text information, but which, through a plurality of separate tokens, of its constituents (with corresponding references to other structural formations) forms a practically cognitive neural network of a certain knowledge area.

**Keywords:** semantic Web, Data Base, integration approach, quantum of knowledge, the basic semantic-syntactic structure.

**THE PROBLEM OF “SEMANTIC WEB – DATA BASE”**

Tim Berners-Lee has been working on creation of the WWW web for over ten years. The main idea is to use a variety of agents for carrying out multiple tasks of users creating separate tracks between the data bits stored on different computers. Practically, this is the distributed system, which enables access to the variety of interrelated documents through the Internet. With the advent of WWW, which to some extent updated the hopes on modeling intellectual capabilities of human, the ideology, presented by the interrelation “Semantic Web - Data Base”, which has been heating the imagination of many researchers and generations up until now, has been created. The first part is related to the problems of perception and “understanding” of a message, while the second deals with generation and utilization of basic knowledge.

In the general case, those are the components of human cognitive potential, which embrace the whole communication process, considering all its constituents:

perception of a message, understanding its content, making a decision, verbalization of the decision, etc., which in general are connected with the first part of the interrelation – Semantic Web. For its time it was a revolutionary step, which dramatically influenced the formation of the whole cluster of information technologies, oriented on modeling human speech activity. However, it is worth mentioning that this aspiring program stimulates the search and development of certain directions of modeling the speech activity, and at the same time, it is still pretty far from cognitive abilities of alive neurosubstance.

The second component – “Data Base” as a part of human cognitive potential represented by global text information files, is practically formed in computer networks, while the first component – “Semantic Web”, responsible for “understanding” the message, and is still on the stage of initial search. Nevertheless, Tim Berners-Lee virtually suggested that realization of this ambitious intellectual project might happen in next twenty years. However, time is passing and the number of unresolved questions does not decrease, it even seems to increase, especially regarding semantics.

In the wake of the euphoria of solving these ambitious perspectives, entire groups, associations, institutes have been formed; presentations and polls have been conducted, etc. The results of surveys conducted by a group of researchers at Elon University (Janna Quitney Anderson — Pew Research Center's) and the American Life Project (May 4, 2010) are considered to be quite interesting. The main result is quite presentable, though cautious. The results of the survey conducted by the research centers on this project are quite revealing.

895 respondents were selected; they would have spoken about the possibility of realizing this project within certain timeframes. Evaluation of the results are the following (based on [1]):

- about 47% of respondents expressed moderate expectations regarding the possibility of realizing the project: “The project would not be as efficient as expected and the average users would not feel substantial difference when the project is realized” – this is the point of professionals who understand the complexity of the issue;
- almost 41% expressed hope that “by 2020 Semantic Web would succeed and would be able to provide better service for average user” — this is the point of those users who are quite unaware of situation and satisfied with the current condition of the Internet;
- the rest (12 %) did not express anything regarding the evaluation of this project.

Thus we can see that the society in general especially the professionals has taken a rather reserved view of the project and further development is nothing more than unfounded illusions regarding future perspectives. Of course, the problem posed and voiced by Tim Berners-Lee refers to one of the greatest challenges – modeling human speech activity, which is mainly realized by our neurosubstance.

We see that the problem of finding information in the gigantic repositories of data stored on electronic media is more or less solved, and partially satisfies the users who have extracted information either through libraries or in dusty repositories, fluttering innumerable folios in the pre-computer era. Yet, in our opinion, the problem of meaningful processing of information remains a priority for humanity.

Can we shift this creative function on to the shoulders of the computer? Naturally, the question arises, where does the problem lie, in which direction to go, and in general — are these issues solvable?

Practically, this program (to our mind) is intended for modeling the processes of using (understanding) of natural linguistic information, which in general quite often is “illogical and somehow playful and mistaken” and oriented on cognitive potential of interlocutor. The last remark is that current natural linguistic technologies are restricted to using them as key words. The key point is that our language is still not analyzed enough; there are still a lot of problem and uncertain issues. Practically, this is the main argument regarding the possibility of formalization of natural linguistic technologies. Obviously, various questions arise, such as what is the reason, why our knowledge about language are so poorly formalized, why, ..., why?

The Semantic Web — Data Base program attracts more and more attention to the problems of speech modeling, on the one hand, due to the wider use of computer technology for applied linguistic tasks, and on the other hand, as it defines a new vector for the development and use of information technologies in the field of modeling information processes in our neurosubstance. That means, we already face the problem of modeling quite complex psychological functions that could not be implemented at previous stages of the development of information technology. First of all, it concerns the modeling of speech activity.

## **INDIVIDUAL LANGUAGE SYSTEM**

In the context of analyzing the interrelation “Semantic Web — Data Base” it is worth addressing to the accomplishments of remarkable Kyiver Lev Shcherba, who since his birth (since being six months old) has lived in Kyiv with his parents, graduated from the lyceum, studied on the first course of Kyiv university. Further, he proceeded studying in St. Petersburg, where under the supervision of Jan Baudouin de Courtenay he was conquering the highest ranks of European linguistic science. Shcherba’s collection of works “Language system and speech activity” and the work “On a three-fold aspect of linguistic phenomena and on experiments in linguistics” in particular are consider the highlight of his linguistic experience and future foresight [2]).

The main scientific work of L. Shcherba is the concept of individual language system (ILS) as a combination of linguistic processor (LP), responsible for structural and functional level of message organization, and basic knowledge (BK), where the complete cognitive potential of a subject presented on language level is stored and accumulated. The key distinction of human “individual language system” and computer models is that nowadays in information technologies all the cognitive potential (basic knowledge) is still presented on language level as fragmented variety of text information files, while in human system this potential is integrated to a single medium, presented by our neurosubstance. Nevertheless, we are fully aware of certain structural identity between Tim Berners-Lee’s method and Shcherba’s concept of ILS, where linguistic processor, responsible for structural and functional level of linguistic organization and provides the function of “understanding” of a message, and basic knowledge is responsible for the whole human cognitive potential. Note that the ISL concept was introduced by L. Shcherba back in 1927, i.e. almost a hundred years before there appeared a chance

to shift our linguistic competence on to the shoulders of computers. Well, let's use this ILS ideology as a combination of LP and BK for solving current problems of modeling human speech activity, create their corresponding models  $M_1$ ,  $M_2$ , and lock them out on each other with the help bilateral connections (fig. 1).

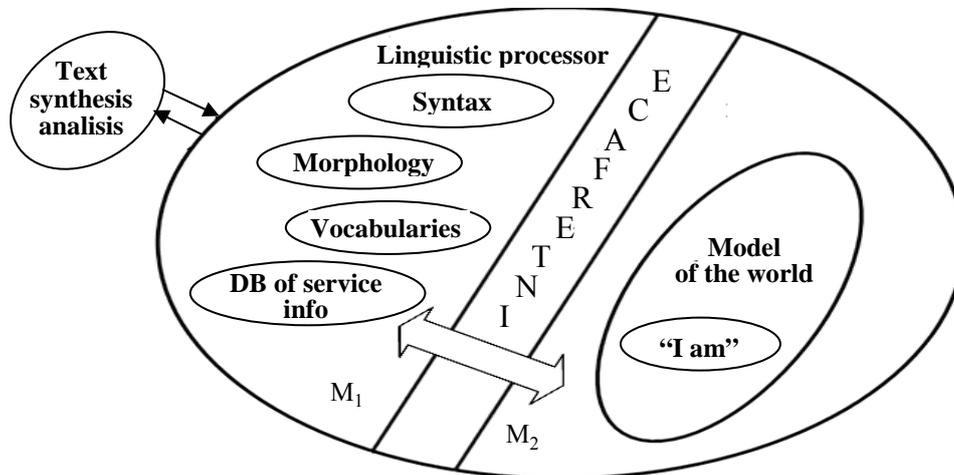


Fig.1. Individual language system: LP — linguistic processor; BK — natural linguistic basic knowledge

In general, the LP is responsible for identifying the structural and functional level of the organization of a separate message, while the basic knowledge (BP) is considered a repository of all cognitive potential, taking into account the structural and functional level of the linguistic organization, but this time it is the cognitive potential of a particular individual. We emphasize once more that there is a difference between human cognitive potentials and computer networks: at present, cognitive potential is represented in natural linguistic information technologies (NLITs) only by significant volumes of text (or speech) information, while human cognitive potential comprises also the information from all sensory levels (sight, hearing, touch, taste, smell).

When, apparently, everything is so transparent, the question arises, why aren't there no technologies for processing natural language information up to this time? The reason is that (on the one hand) the language is such a complex and multifaceted object for exploring, and especially for research, that it integrates practically all directions of analysis of this phenomenon, starting with biology, psychology, neurophysiology, philosophy, cybernetics, and other interdisciplinary areas, (and on the other hand), language is still not sufficiently structured for formal analysis and modeling.

The first author's attempt to integrate diversified knowledge on linguistic organization was in 1998, when on the occasion of century of establishing Kyiv Polytechnic Institute (KPI) the author's study guides "Language Architecture" [3] was published. In this guides, from the perspective of statistical average level of analysis of current state of structural linguistic organization (quite reflexively), the notion of "basic semantic and syntactic structure" "(BSSS)" as a basic structural element of linguistic organization was introduced. However, the research was some kind of generalization of that time knowledge regarding structural level of linguistic organization, and did not reveal anything new in the sphere of linguistics.

The qualitative leap of author's approach to structural and functional level of linguistic organization is based on many years of teaching the course “Sensory Systems” at the Department of Technical Cybernetics at Igor Sikorsky Kyiv Polytechnic Institute (KPI), during which the structural and functional organization of all systems of sensation, including the stages of perception and processing of arbitrary information, were carefully considered. The significant moment was the familiarization with the works of Semir Zeki, concerning the structural and functional level of the neuro organization of the visual pathway [4] and the work of J. Hawkins “On Intellect” [5], summarizing the functional identity of all sensory systems. Therefore, thorough analysis of the functional load of sensory systems, taking into account the unresolved problems of linguistics, allowed synthesizing a rather harmonious model of the formation and development of speech activity.

It is important to emphasize the significant stage of the author's plunging into neurophysiology of sensory systems after meeting and fruitful contacts with the academician O. Kryshchal, which stimulated publication of the article “Neurophysiological bases of linguistic organization” in the NASU Reports (upon the recommendation of O. Kryshchal) [6] and small but important author's monograph “From Thought to Knowledge”, published by Ukrainian chronicles in Kyiv in 2008 [7]. In these works the way from perceiving the information of visual level to translating it onto language level has been thoroughly researched. All this allowed making a solid conclusion that the structure of linguistic organization is greatly defined by the level of structural and functional neuroorganisation of visual pathway.

The title page presents a procedure for translating information of a figurative level onto language level in the form of a sequence of individual steps. In general, visual information enters the retina, which is filled with receptors of two types: sticks (in the number of  $130 \cdot 10^6$  located on the periphery of the retina) and cones (in the number of  $6 \cdot 10^6$ , filling the central foveal area, and capable of thorough identification of the color range). When we want to get a detailed look at something from the surrounding environment, we project this particle (with the help of the lens muscles) to the central foveal area. Such fragment of the image will be determined as a situation. The situation is a fragment of visual component of environment, which gets to the central “foveal area” of retina and is processed to its full extent. In each particular situation all its components Obj/Subj, their dynamics Mov, their attributes Attr (Obj/Subj), Attr (Mov) and the extent of those attributes Attr (Attr).

In section 3.4 (Verbalization of visual information), taking into account the results of neurophysiologists and psychologists research, the procedure for processing a particular situation of a figurative level with subsequent translation of its results onto the language level is clearly traced. It is important to note that this is a procedure for processing only one situation, and it is implemented with a frequency of 25-75 Hz, tracking both the static and dynamic characteristics of the components.

Only a part of visual component, which gets on the central foveal area, that is a tenth part of a visual field, is thoroughly examined. However, it is processed quite elaborately: through identifying all objects/subjects, their attributes and the extent of these attributes, as well as identifying dynamic characteristics of all the defined components, with further identification of their attributes and the extent

of the attributes. This is, virtually, the scheme of processing of only one fragment of an image on the retina – a situation. Although, in order to explore the visual field totally we need to examine other situations as well. But on the way of evolution of visual pathway this function is realized in another way through using “sacade”, the system of sporadic scanning of the retina completely in order to find particular components. As we can see this algorithm is quite economical and we don’t have to scan the retina field each time, it is enough to scan only those zones, where the important (for the subject) components are. Thus, we examined the whole path of processing both of certain situation for identification of all its components with defining all objects, subjects, their attributes, extents of attributes, and the complete image, getting to the retina. If it is true, there is still one question — how the process of verbalization of information from sensory level to the language one is carried out.

The first attempt by the author to integrate the accumulated diversified knowledge of the linguistic organization (more precisely, about the stages of its formation and development) was presented in the form Pre-Conference (poster + workshop talk) on summit at the BICA-13 conference in Grandotel, Kyiv, where the integration platform of linguistic organization “Systematic approach to the modeling of speech activity” was presented.

The comprehensive view of such approach was shortly in BICA-14 journal by the article “Back to basics of speech activity” [8]. Practically, it meant the presentation of author’s vision of the problem of linguistic organization to European market. It is important that an extensive axiomatics of speech activity, which turned to be not only possible but also quite important for understanding diversified functioning of speech activity, was presented. The next work of Yu. Kislenko and postgraduate D. Serheev [9] was devoted to the problems of modeling of knowledge base on the mentioned principles.

## **INTEGRATION APPROACH TO THE ANALYSIS OF LINGUISTIC ORGANISATION**

Tim Berners-Lee suggested the Semantic Web – Data Base interrelation about twenty years ago, but the sense of progress in its implementation is still not very visible. Why is this so? On the one hand, this is due to the fact that the problem itself touches upon the issues of modeling complex cognitive processes occurring in our neural circuit, and on the other hand, probably, when the problem is clearly defined and voiced, presumably there are already certain horizons (achievements and certain hopes) to solve it. Let’s try to evaluate the prospects of solving the problem more substantially, taking into account those developments that have already been implemented. In particular, the author’s last work “Personalized cognitive feedback as a powerful lever for accelerated social development” [10] is in some way related to the analysis of the cognitive processes that occur in our neural circuit and determine the progressive accumulation of cognitive potential. It is necessary to clearly identify the platform from which the study begins, in order to clearly understand how and where to go in this direction. Since the problem of “Semantic Web – Data Base” is related to modeling of cognitive processes in our neural circuit, and for their analysis, we have only a speech level available, then we face the problem of analyzing human cognitive activity.

### **Current state of classical linguistics**

The stated research direction is a rather ambitious project, and the author absolutely soberly assesses its complexity. However, here we can trace such a solution: we should only turn to the analysis of the current stage of development of classical linguistics through the evaluation of modern achievements by recognized representatives of this direction. We should only recall that the first qualified study of the structural level of the linguistic organization was presented by "General rational grammar" (the grammar of Port-Royal in 1660) by the authorship of the logician and philosopher Antoine Arno and the grammarian Claude Lancelau [11]. The grammar was formed according to the results of the analysis of the structural level of the linguistic organization of significant volumes of texts, mainly religious. We must pay tribute to the creators of this work, who "felt" and identified the main features of linguistic organization in the form of a dichotomy "simple / complex" sentences, and also identified a certain category of "word combination" for which they did not have a clear definition at that time. Only now, we can assert that such a structure occupies the status of a separate full-fledged sentence, if we assume that the "missed" element of such a structure was just mentioned in the previous sentence and is still activated for a certain time in our neural circuit. This is practically a "saving" scheme of linguistic means, oriented on listener.

The feedbacks from prominent linguists on the current state of the structural level of the linguistic organization are presented in Table 1 from the work of BICA-13. We see that practically the entire linguistic elite is rather restrained in relation to the current state of our knowledge of the linguistic organization. Particular emphasis should be placed on the opinion of L. Astakhova (head of the department of the German language of the Dnipropetrovsk State University), who at one time clearly expressed her attitude to this problem: "Linguistic community has long been ripe for rejecting the existing theories of a sentence; the object of syntactic research is as well unknown" [12]. That is, these are key, painful issues of classical linguistics, which are still waiting to be solved.

The revelation of B. Horodetskyi regarding the current state of linguistics became an impetus for the formation of a coherent, holistic picture of the linguistic organization. This is a person who has edited the collection "New in Foreign Linguistics" for more than 10 years and kept the pulse on key pan-European linguistic problems. His verdict regarding the prospect of a language organization study looks very categorical, but rather constructive: "Many troubles in linguistics are due to the fact that language is still considered a form of reflection of "thought" rather than a scheme for the organization and presentation of knowledge. So, from here we have a powerful conclusion: in the realm of classical linguistics it is first necessary to deviate from the concept of "thought".

The revelation of B. Gorodetsky regarding the current state of linguistics became an impulse for formation of coherent consistent picture of the linguistic organization. This is a person who has been editing the collection "New in Foreign Linguistics" for more than 10 years and kept the pulse on key problems of European linguistics. His verdict on the perspectives of linguistic organization research looks rather categorical, but still constructive: "A lot of linguistic problems are related to the fact that a language is still considered as a form of conveying a thought rather than a scheme of organization and transferring the knowledge [13].

Therefore, hence we have a powerful conclusion that in the realm of classical linguistics, it is first necessary to deviate from the concept of “thought” and master and use the productive platform of “knowledge”.

Thus, for us, the generalized assessment of the current state of classical linguistics and the way out of this critical state, when the concept of “knowledge” is adopted as the basis of the research, instead of the “thought”, suggested by B. Gorodetskyi becomes the directional sign. This is a very appropriate suggestion, since current linguistics still defines language as the scheme of reproduction of “thought”, which is a standard scholastic situation, when one (not clearly defined concept) is interpreted through another, which is also not so well defined. However, the question remains – what is defined as “knowledge”, how is it formed for a human and how is it used?

**Table 1.** Current state of classical linguistics

<b>Authors</b>	<b>Unresolved problems of classical linguistics</b>
Whitney W., F. de Saussure	Language is a system of linguistic units of different levels with no logical connection between them.
Piotrowski R. (St. Petersburg)	Linguistics is not a theoretical science built on experimental data, it is a exploitative science based on brief samples [20]
Astakhova L. Dnipropetrovsk state university)	Linguistic community has long been ripe for rejecting the existing theories of a sentence; the object of syntactic research is as well unknown [12]
Grammar – 70	All current knowledge of classical linguistics cannot be considered from the perspective of a whole system [19]
Grammar Port- Royal	The problem of word combination still has not been resolved [11]
Gorodetsky B.	A lot of linguistic problems are related to the fact that a language is still considered as a form of conveying a thought rather than a scheme of organization and transferring the knowledge [13]

The answer is formed naturally and fairly transparently if one traces the entire chain of perception of information by the sensory system and the successive stages of its processing. Knowledge is the end product of a very powerful intellectual activity of a person, covering the stages of perception and processing of information by the sensory system, with subsequent identification of the obtained results in the cerebral cortex with the subsequent possible translation it onto the language level. This definition embraces the entire chain of transformations from figurative to language level, but further, we will explore the entire sequence of stages more thoroughly.

Therefore, it is important to observe what is happening on the way of perception, formation and use of linguistic information, but this is an appeal to neurophysiology that is rapidly developing in the modern world and provides answers to many questions. Figure 2 presents a difference in the approaches to the process of synthesis of linguistic material. Classical linguistics was formed on a set of finished texts, not plunging into solving problems regarding the nature of formation and processing of a linguistic message. By this time, it adheres to the idea that “we perceive a human through language”. On the other hand, when we stand in the position that “text is the end product of a rather powerful intellectual activity of a person” which necessarily includes the stages of taking into account the dependencies of the perception of the environment by the sensory system (ranging

from retinal functions that determine all components of a particular situation and ending with the functional load of the visual cortex, which defines all their attributes), we can find out what knowledge is, how it is formed and used.

Integration approach to structural level of linguistic organization covers the results of recent researches of linguistic organization in various joint spheres, such as biology (Haeckels law); neurophysiology of visual pathway (Hubel, Wiesel, Semir Zeki, Jeff Hawkins); psychology of perceiving the environment (defining a situation); philosophy (the structure of a message in the triunity of time, space and action); cybernetics (ontogenetic parallels law) etc.

### Defining the category of “knowledge”

We can present a certain sequence of language message formation taking into account the following steps. Our sight, like other sensor systems (Semir Zeki, [4], J. Hawkins [5]), works discretely at 25–75 Hz frequencies. So, for the “knowledge quantum” that is perceived and processed by a human, it is worth taking a separate discrete (separate frame) of visual perception of the environment when it is assumed that human perceives a lion’s share (80-90 percent) of all the information that comes in general through sensory system through the visual system.

Speech activity is one of the most significant functions of human society and, probably, one of the least explored areas of society’s existence and functioning. However, to date much knowledge in many areas of research of various spheres of its operation has been received and accumulated. For that reason, while exploring this issue it is worth “integrating” together all the knowledge received in different spheres.

### Speech activity as a research object

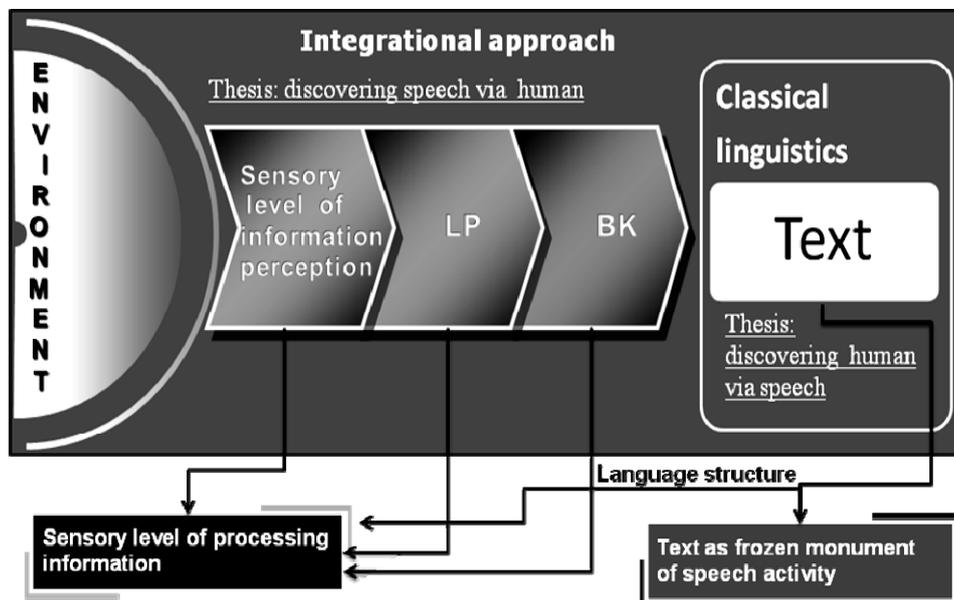


Fig. 2. Integration approach to creation of linguistic organization: Defining principle — “we perceive the language through human”

Despite author’s “solid experience” in the domain of the linguistic organization (generally about 40 years, and over fifty publications), he managed in a cer-

tain way to form a rather clear scheme for the establishment, formation and use of speech activity.

Practically, halfway comprehensive and uncontroversial picture regarding linguistic organization has been formed over the last decade, when the author gradually started overcoming the problems of speech activity taking into consideration current researches in various related fields. Relation of integration approach and classical linguistics is well illustrated by the slides from author's work for the Conference BICA-13 in the form Pre-Conference. On the slides, the comparative characteristics of integration approach regarding the long-established accomplishments of modern linguistics. Classical linguistics adheres to the idea that "we perceive a human through the language", while the integration approach changes the vector of research for the opposite – "we perceive language through human"; that means we are going to consider the way how human perceive the environment, how process, generates, and translates it on language level, as well as how human uses it.

Such approach practically settles a lot of problems of classical approach to structural level of linguistic organization since it traces the whole path of perceiving and processing of certain knowledge quantum through the visual pathway. Thus, we are considering a "knowledge quantum" of visual level as a particular "frame", which we further define as "situation". Taking into account the researches of neurophysiologists of visual pathway, we can define this notion as follows: "situation is a fragment of visual component of environment which gets to the central foveal area of retina and is processed to its full extent".

On average, to the "central foveal area" of retina (according to psychologists) only ten percent of the information perceived by the retina gets. However, when about a hundred objects / subjects fall into the retina, in general, we see that no more than ten components gets to the central foveal area, which are then processed fully, and after that, they can be translated into the language level. Surprisingly this number totally coincides with the statistical data of linguists, who state that the length of a simple sentence does not exceed seven plus or minus two components. The general scheme of processing information through the visual pathway was well illustrated in author's monograph "From thought to knowledge" [7], in which the schematic overview of the sequence of stages on the way to realizing the interrelation "Reality–Text" was presented (fig. 3).

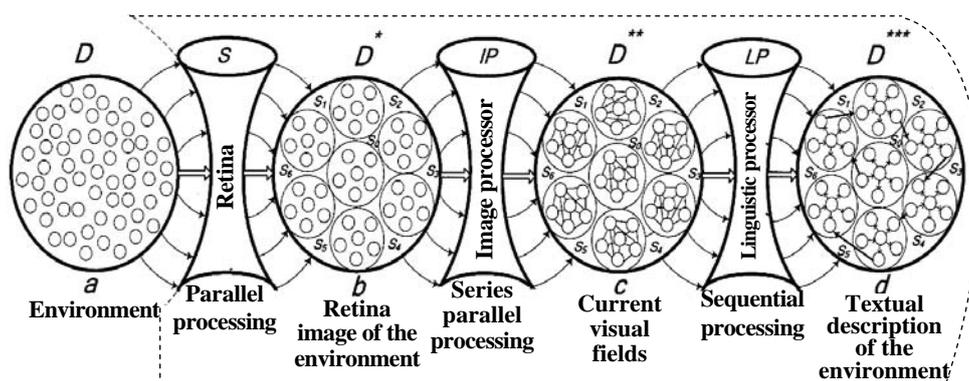


Fig. 3. The stages of processing the environment in the direction "Reality–Text": *D* — the environment getting to the retina; *S* — the variety of situations, that could be processed sequentially; *D\** — identifying situations; *IP* — image processing; *D\*\** — image level of processing the situations; *LP* — linguistic processor; *D\*\*\** — symbol level of presenting the environment

It is important to mention the perception of the environment through visual pathway, hence processing of received information, is conducted with a frequency of 25–75Hz. Such scheme ensures clear identification of not only static, but also dynamic objects.

### **Stages of formation of the “Basic semantic and syntactic structure”**

How does the individual situation is processed by the visual pathway? — Careful researches of neurophysiologists give the answer to this question. According to Semir Zeki [4], who has been summarizing the results of studies of the visual pathway during the last half of the century, we have:

**experimentally-confirmed** the presence of ensembles of neurons (hundreds / thousands), which select all objects / subjects (Obj / Subj) from a separate situation (the third level of retina);

**experimentally-confirmed** the presence of ensembles of neurons (hundreds / thousands), which, select dynamic components with movement identification (Mov) from a separate situation (fourth, fifth levels);

**experimentally-confirmed** the presence of ensembles of neurons in the visual cortex of the brain (hundreds / thousands), which identify their attributes Attr (Obj), Attr (Subj), Attr (Mov) for all found components (Obj / Subj) and their dynamical characteristics (Mov);

**experimentally-confirmed** the presence of ensembles of neurons in the cortex of the brain (hundreds / thousands), which also determine the extent Attr (Attr) for the found attributes.

So, neurophysiologists gave us a clear answer to how the process of treating the visual pathway of a particular situation is performed: with defining all of its components, identification of dynamic characteristics, as well as with the clear ability to differentiate all the diversity of attributes. Therefore, after processing a separate quantum of knowledge - situation, a person should only “mark” the components of the situation determined by the visual pathway with corresponding language labels. But this is a very difficult, hard and long process of language formation in human society.

The answer to how is the situation translated into the linguistic level was given by A.N. Gvozdev, who had comprehensively and very carefully been traced the peculiarities of the process of forming the linguistic system of his son Eugene for seven years. He did what often many parents try to do, but he was a person with higher linguistic qualifications who systematically and purposefully compared every new step of mastering the language with a specific environment. The results of the research were published in the academic edition in 1949 with dedication to his son [14]. These observations show a clear (Fig. 4) staged scheme of formation of the linguistic system of a child, which ends at about three years.

At this age, the child finishes the process of forming the main structural element of linguistic organization — the basic semantic and syntactic structure (BSSS), as the main scheme of verbalization of a particular situation. Researchers of child's language even claim that at this period the child becomes a “professor of linguistics” at the level of processing the household language level. But this is a very difficult, important and responsible period both for parents and for the

child, when you need to calmly and effortlessly compare and correct each component of every real situation (subject, object, action, time, space, reason, condition, consequence etc.) with the language equivalents of their verbalization. We should mention Masaru Ibuka and his bestseller “Kindergarten Is Too Late” [15], which excites all parents and future mothers. Why is it “late”? – Because, by this time child’s neural network is the most plastic and effectively “sucks” (like a sponge) all information about the outside world. Over time, this ability becomes less effective, and experts argue that if a child has not mastered the language by five years, then, in practice, he is not able to become a full member of the society. Fig. 4 represents this level in stages (1-5), when the child learns the procedure of verbalization of a particular situation; the last stage (stage 6) represents the process of synthesizing the message at the polypredicative level.

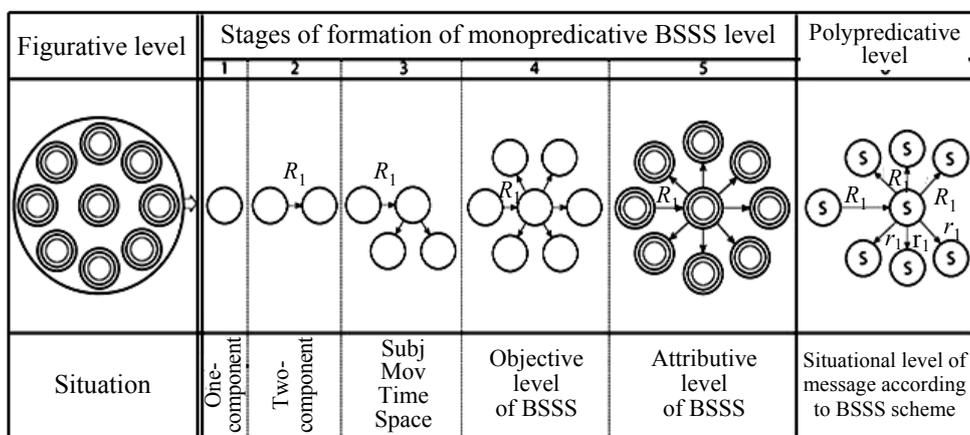


Fig. 4. Staged scheme of formation of the linguistic system of a person

The sequence of these stages, and considering their functional load, a clear organization of the verbalization scheme of a particular situation emerges in the form of a standard procedure which is mastered by a child at the age of 2,5-3 years. Basically (according to the definition) — the structure of the BSSS is a two-component, monopredicative scheme describing any situation from the real or virtual world, all components of which are updated at the attributive level. Practically, the Fig. 4 presents the sequence of stages of the child’s learning the message structure of any complexity.

The functional organization of the BSSS structure (monopredicative level) is mastered quite easily and on time, while the polydynamic scheme of message organization is learned later and harder; quite confidently the child masters it at the age of ten or twelve years. The peculiarity of formation of the linguistic order of the polypredicative level is that the system of functional connections remains adequate to the BSSS with certain transformations, and a whole structure of the BSSS may be put in the place of each component and so on, proceeding to the level of reproduction of the recursive scheme of the message organization. In general, the way of formation of this structure is carefully analyzed by the work “To the origins of speech activity” [8], and the structure of the BSSS itself is presented at Fig. 5.

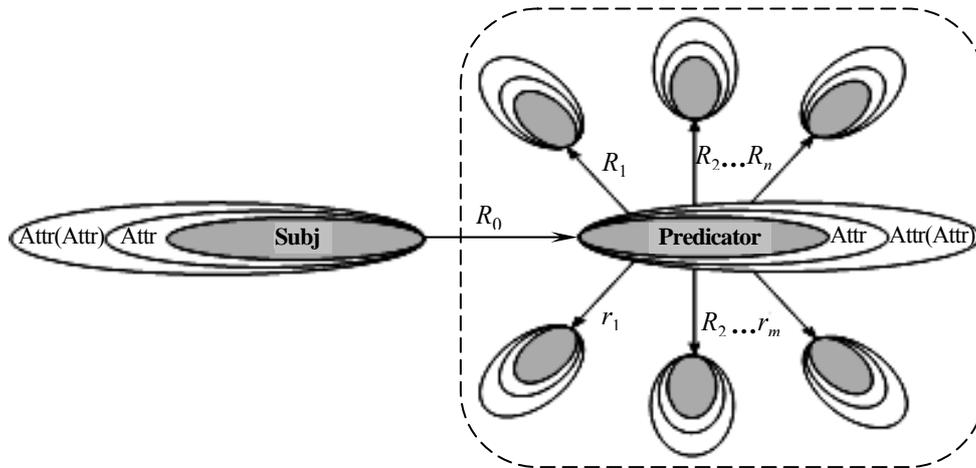


Fig. 5. Basic semantic and syntactic structure: Subj — subject; Pred — predicate;  $R_0$  — main relations (mother-predicate);  $R_1 \dots R_n$  — predicative relation;  $r_1 \dots r_m$  — syntagmatic relations; Predicator — core of the predicate

### Forming messages at the polypredicative level

The BSSS structure (Subj – Pred) reflects only to the monopredicative level of organization of a message, whereas the message can be formed generally on the predicative level, which combines several BSSS structures. Generally, this will be a structure similar to the BSSS structure (Fig. 3), where the whole structure of the BSSS may be put in the place of a separate component. At the same time, only the system of relations, which connects separate situations, changes. The issue of forming messages at the polypredicative level is well researched by Yu. Kyslenko, A. Khimicha “The structural and functional level of organization of the linguistic processor for informational natural and linguistic technologies” [16].

The fundamental difference between the formation of messages of the polypredicative level is the usage of another semantic load as components of a separate structure, as well as systems of relations in comparison with the BSSS structure. When the situational relations of the BSSS structure determine actual coordinates of the situation in the surrounding world or environment (time, space, reason, condition, consequence, etc.) then, in case of the polypredicative type, these coordinates are formed in relation to another situation, using a relative relation type: “where there is  $S_1$ , then  $S_2$  occurs”, “when  $S_1$  is implemented, then  $S_2$  will be”, and so on. In practice, the polypredicative message has a strict organization and is clearly identified by a linguistic processor. Transformation of predicative and situational relations is clearly analyzed in the following publication. Generally, the scheme of forming a message of the polypredicative level can relate not only to the system of relations of the BSSS structure, but also affect the relation at the attributive level (let's recall a familiar example – “This is the house that Jack built. This is the malt that lay in the house that Jack built...”). It is important that the structure of the polypredicative message is also clearly identified by the linguistic processor.

It should be once again mentioned that if an environment is perceived by a person as a sequence of individual situations, each of which is processed by the visual system under the same scheme, then the structural organization of the lan-

guage, almost regardless of nationality, color of the color of skin etc. should be processed under a standard scheme. Of course, languages differ significantly in terms of lexical composition, schemes of formation of polypredicative level, but their structural and functional organization have many common issues.

### **The final part to the integration approach**

When the integration approach to the linguistic organization determines that the linguistic organization is derived from the structural and functional level of the neural organization of the visual pathway of a person and does not depend on the nationality, color of skin, place of residence, etc., but is determined only by the human genome, then we come to a significant conclusion that at the neural level each separate situation should be processed in the same way - with defining all components of the environment OBJ / SUBJ, their movement Mov, with further identification of all their attributes ATTR (Obj / Subj), ATTR (Mov) and also the extend of the specified attributes ATTR (ATTR). That is, the study of a situation at the structural and functional level of the neural organization of the visual pathway occurs in the same way, regardless of nationality, place of residence, color of skin, etc. The difference is in the linguistic organization of certain components (language labels), which identify the results of processing of a situation, in the levels of lexical identification of certain environmental components, in the order of their use, and also at the level of formation of mono/poly- predicative structures. Actually, this level finishes the analysis of the structural level of a particular situation, which is translated into the linguistic level by the BSSS structure.

Therefore, we have a significant conclusion that all the cognitive potential received, generated and accumulated by humanity, at the linguistic level (regardless of the language), will be structured by almost the same way based on the sequence of the BSSS structures (see Fig. 4), despite certain differences in their organization at lexical level, and at the text formation level in general. So, we can now talk about the organization and accumulation of a global cognitive knowledge base with free access thereto. Certain developments in this field have already been made.

## **PROSPECTS OF FORMATION OF SEMANTIC WEB**

### **Information load of Semantic Web**

The integration approach gave us a coherent system of the structural level of the linguistic organization, which is based on the standard BSSS structure. This is the syntax of the language organization of an arbitrary message. So, naturally, the question arises - what is semantics, and how is it formed? In our opinion, semantics (semantics of a message) is a system of binary relations between elements of the BSSS structure, the order and direction of which are determined by the scheme presented at Fig. 4.

If we take the BSSS structure as a basis and remove all the relations identified by the system of relations:  $R_0$  is the main relation “mother-predicate”, which connects Subj and Predicat together;  $R_1, \dots, R_m$  — relations that clearly identify all the predicative components of the predicate;  $r_1, \dots, r_m$  — situational relations that determine the coordinates of a particular situation (time, space, reason, conditions, consequence of actualization of a particular situation), what does remain? Apparently, there is an unordered dump of lexeme with no semantic loading at the entire message.

It is worth noting, however, that the semantic load of all components of a situation is gradually realized, shaped and generalized by the child in the way of mastering the linguistic organization. That is, the process of mastering the language also means the formation of a semantic message system at a monopredicative (age up to 3–5 years) and a polypredicative levels (age from about five to twelve years).

Still, the question remains of how does the recipient identify these attitudes at the stage of perceiving the text? The answer will also likely be clear - either by using a certain lexical load of components that determine, for example, time (hours, minutes, seconds), stages of the day (morning, day, night, etc.), etc., or through the link to other situations, which were uniquely identified earlier.

This small excursion to the semantic relations of the BSSS structure draws the attention of the readers to the problem of formation and usage ("understanding") of semantic relations to search for relevant information; moreover, they are clearly identified either in separate word forms or have a relative reference scheme to other quantifies of knowledge. Well, we can proceed with analyzing this problem.

So, there is the problem of synthesizing the Semantic Web in such a way that, in part, it was possible to identify semantic links from textual information. We perfectly (and almost automatically) identify these relations, because during the long evolutionary way, we have worked out such schemes for identification of semantic relations, which ensure adequate perception and reproduction of text at the stages of analysis and synthesis of the message. Again, please remember that the semantic load of a separate BSSS structure is determined by a system of functional (semantic) relations of its constituents. So, we can predict the Semantic Web for one component of the message presented by the BSSS structure. One should only consider that for our language, the order of BSSS formation can be free.

We live in the period of computerization, we often quantify our thoughts with "bits" or "bytes", but not always use "knowledge". The language, the writing, is a great achievement of humanity, it is something that distinguishes man from animals. We have learned to record our thoughts, our experience for the term of preservation, which far exceeds the life cycle of a person.

It is time to pass it to your computer, in other words - to put knowledge to the machine, presented by a plurality of texts accumulated and saved by humanity. We can confidently identify the temporal and spatial relations, reason, condition, consequence, etc. in the texts. This knowledge should be "put" into the memory of the computer in the form of a linguistic processor (LP) and used. These are almost simple problems that can help programs "understand" the text by identifying individual functional relations between the words. This is the main meaning and purpose of the computer – not to lose the accumulated knowledge and learn to understand and use it.

Again, we return to the problems of modeling the Individual Language System (ILS), which clearly separates two components: the linguistic processor (LP) and the basic knowledge (BK). Let's carefully form the BK, where, at the textual level, there will be formed all the major accumulated cognitive potential and "gradually" we will form a linguistic LP processor to teach the "foolish" machine to "understand" the text, gradually putting separate bricks of knowledge thereto, presented by the plurality of interactions of the same BSSS structures. It is important to capture the semantic load of the individual BSSS functional relations (to be

a reason, to be time, to be a place, etc.). The structure of such relations can be easily and unambiguously identified by a linguistic processor. All components of the situation - Obj, Subj, predicate, situational relations (time, space, reason, consequence, conditions) - are easily identifiable and work for us).

Let's remember A.S. Narinjani and his school, which deals with problems of "time", "space", etc., Let's also recall N. Leontiev, who argued that the texts always have all the information necessary for a thorough analysis of the message and the use of this information. This issue is partly covered by the publication of the author "Structural and functional level of organization of the linguistic processor", which actually analyzes the means of identifying situational relations for the mono/poly- predicative levels of linguistic communication organization [17]. These works also presented a virtual experiment regarding the possibility of searching the Internet for a "full text". The main components here are components of Subj-Pred, situational relations (time, space, movement), predicative relations.

It is important to emphasize that the system of relations of the BSSS structure at the mono/poly- predicative levels cements the set of components of the "lexemes" into a monolithic knowledge quantum, which determines the "semantic load" of a message. That is why special attention should be paid to connecting the set of unsystematic lexemes to a semantically executed message. What do we have now?

### **Comparative analysis of various schemes of formation of "Semantic Web"**

The integration approach to the structural level of the linguistic organization, presented in Section 3, gives a clear vision of the linguistic organization both at the monoprostand level, which is confined exclusively to the BSSS structure, and at the polypredicative level, covering a plurality of such structures, based on general level of the recursive scheme of the organization of the message. However, the basis of the structural level of the linguistic organization has always been the basic semantic and syntactic structure — BSSS.

In practice, the basis of the entire structural level of the linguistic organization is the basic structure, the semantic framework of which is determined by the system of relations of the BSSS structure both situational and predicative, presented at Fig. 4. Below is the list these relations:

- ratio  $R_0$  — the main relation of the structure that identifies the connection "Subject-Predicate" — (Subj-Mov);
- the predicate itself with the "*Predicator*" core, which covers the verb (or its transformations according to the schemes of the adjectives or adverbs, based on two groups of relations - situational and predicative;
- situational relations  $r_1, r_2, \dots, r_n$  — determine time, space, reason, conditions, consequences of action. . . in the environment;
- the predictive relationships  $R_1, R_2, \dots, R_m$   $R_1, R_2, \dots, R_m$  determine other entities / objects that are involved in the actualization of the action identified by the predictor Pred.

It is important to emphasize: the total number of elements that are simultaneously defined by one situation does not exceed from seven to ten. This is the average number of components that can simultaneously be processed by our visual pathway. In case of a polypredicative message, this number will depend on

a plurality of sequentially reproduced situations. It is also worth considering the number of attribute relations of all involved objects / subjects, when they are identified as: Attr (Obj / Subj), Attr (Mov), Attr (Attr). This is a system of components that uniquely identify the attributive level of an environmental situation. The person is not able to perceive the situation in more details on a figurative level and identify it, respectively, at the language one.

So, when we have determined the scheme of perception of an environmental situation, and transform it into linguistic level, we can now correctly put the question about the possibility and effectiveness of using this system of relations to search for information in powerful natural and linguistics (NL) information repositories.

### **Modern ways of presentation of Semantic Web**

We have just studied how adequately, fully and thoroughly a person may perceive environmental information, considering the semantic load of all components of a situation and translate it into the language level. Here the process of the recipient is actualized in the direction of "Reality – Text", whereas for users (listeners / readers) it is important to reproduce this dependence in the opposite direction – "Text – Knowledge".

The problem for the readers to apply to the gaining potential of knowledge presented at the computer level, is solved so far, considering the artificial languages of NL texts. Such first attempt was implemented in the form of SGML (Standart Generalized Markup Language), approved in the 80's when a certain list of instructions (tags) was created to reproduce the text structure. HTML is a simplified version of SGML structure that defined a certain set of Tags, their attributes, and the internal structure of the text defined by the DTD rules for certain types of document. However, it is important to emphasize that, in fact, the texts themselves, as cognitive enhancements of our neural network, are by no means connected with the Tags used for formal presentation of texts; nor does it consider the semantic relations between Tags, the number of which is limited. Considering the situation, the specialists have already expressed the idea that HTML today does not fully meet the requirements of both the developers themselves and the users. The proposed version of XML (according to the design of the developers) itself should "synthesize" a specific set of tags by text; but this means that it must have its own "intelligence" in all areas of knowledge. Currently there is no answers of how to formalize this semantic problem!

Given these unresolved issues, the specialists state that: the evolution of data structuring systems under the Obj-Attr scheme gradually brings researchers to the complication of tags, which ultimately requires the ability to analyze the entire structured level of text information.

Because of the analysis of the current state of the problems of searching information on the Internet, considering the speed of computers, the possibility of presentation of information in distributed systems there raises a question – should one refer to a "full text" search? Certain proposals have already been stated [18], although the author has not yet found this proposal.

The author is not an expert in computer technology but has achieved a good level of generalization of the structural level of the linguistic organization, which is derived from the structural and functional level of the neural organization of the visual pathway, and therefore - will be the same for all languages at the structural

and functional level of their organization. In connection with this, there is a suggestion: to transform the search system according to the BSSS standards, as a generalized scheme of synthesis, analysis and presentation of a linguistic organization regardless of a language.

Let's consider modern schemes of indexing the linguistic material, which are used for searching in modern technologies.

### Using triplestore

The indicated ratio covers a large amount of developments related to the knowledge and modeling of two diverse ways of modeling the linguistic activity of a person. The first component is Semantic Web, which is related to modeling the process of “understanding” the text information, while the second — the Data Base is associated with modeling the processes of accumulation and use of human cognitive potential. So, it's worth differentiating these, interconnected problems, to thoroughly analyze their features.

The concept Semantic Web was introduced by Sir Timothy Berners-Lee in the early 2000's. Conceptually, this is the development of the Internet in the direction of presentation of knowledge (again this uncertain term “knowledge”!). By this time, Internet resources present this concept by unstructured WEB3.0 xb WEB of DATA content, which is accompanied by certain metainformation. A similar representation allows to create (intelligent / semantic) information systems that are oriented towards “understanding” the content. The ideology of Semantic Web uses many formalisms fixed in standards.

The main scheme of information representation in Semantic Web is a triple, which updates the combination of subject, predicate and object. The structure of this representation is presented in Table 2.

**Table 2**

Subj	Predicate	Object
my apartment	has	my computer
my apartment	has	my beg
my apartment	is in	Philadelphia

The concept of triple is so generalized that with the help of an unlimited number of triples we can describe anything. Next, in Semantic Web, as well as in general, in Computer Sciences it has been agreed to share data and data model (metadata). In this case, the model / metadata are implemented in the form of the ratio of ontology / taxonomy using OWL / RDFS, and the actual data are realized in the form of RDF-triples.

The peculiarity of Semantic Web is that there is no clearly defined structure, as in the world of relational data, where the types of relations between the elements are recorded and stored in accordance with the model. The structure is quite complicated. In 2000, many different schemes of triplestores were created. Triplestore is the foundation of Semantic Web.

### MIVAR approach to forming the “Semantic Web – Data Base” model

Another approach to solving the problem of “Semantic Web - Data Base” was proposed by O.O. Varlamov (Moscow), where for the presentation of text information and search implementation, there was used another triade – “mivar” as a

ratio “If ... $S_1$ ... , then ... $S_1$ ... ”. This is, in practice, a model of metadata, which, according to the author's idea, should solve the problems of the relation of “Semantic Web – Data Base” in both directions:

- on one hand, it is used to create, accumulate and use the global database and the rules of their use based on adaptive discrete, mivar informative space based on the triade “object, property, relationship”. It was assumed that the technology could be used both for accumulation of knowledge (Data Base formation) and for search of information (Semantic Web);
- on the other hand, the mivar technology was developed to form a logical conclusion using the ratio “If ... $S_1$ ... then ...  $S_n$  ...”, where  $S_1, S_n$  are related situations connected by certain logical connections.

Somewhere at one of the conferences of the Seminar “Artificial Intelligence” in Katsiveli (Crimea) the author managed to hear to the report O.O. Varlamov. At that time, according to the speaker, there was accumulated a mivar base, which included several million mivaras. A positive point is the use of logical conclusion about the plurality if mivars. Such base can also be presented by the triples “If ... $S_1$ ... , then ...  $S_n$  ...”,

However, despite the principled opportunity to present the environment through such triplets, the procedure itself turned out to be quite complicated and unusual for users of the natural language. On the other hand, real text turned into a meta-text that is not badly perceived by the computer, but people who were engaged in this unnatural transformation of the text, had a lot of problems.

### **BSSS-option**

It's time to present the problem of searching the information under the “full text”, which is based on the basic semantic and syntactic structure of the BSSS, as the main and standard structure of the presentation of text information. So, it is worth to compare the existing Semantic Web presentation schemes with the system of relations of the BSSS structure (Fig. 5).

For the system “triples – store” we have a set of standard triples “Subj – Pred – Obj”, through the sequence of which (according to users) it is possible to reproduce (present in a certain way) an arbitrary text.

For the mivar approach, on the one hand, the standard scheme of the triples “Subj – Pred – Obj” is used, and for forming the logical conclusion, there is used the sequence of triples of another kind “If ... $S_1$ ... then ...  $S_n$  ...” is used.

However, in both cases, experts who perform the procedure of indexing the textual information under the unnatural scheme of text processing, express certain claims to customers.

Let's now look at the system of relations of the BSSS structure, which we have mastered since childhood and perceive and use in a completely natural way. The integration approach covers a system of simple binary relations:

Subj – Pred	Subj – Attr(Subj)	Attr(Subj) – Attr(Attr)
Pred – Obj1	Obj1 – Attr(Obj1)	Attr(Obj1) – Attr(Attr)
Pred – Obj2	Obj2 – Attr(Obj2)	Attr(Obj2) – Attr(Attr)
Pred – Objn	Objn – Attr(Obj2)	Attr(Objn) – Attr(Attr)

This is almost the system of relations of a separate structure of our native language, which we use daily - it is part of our language: (noun, verb, adjective,

adverb), the combination of which form the standard BSSS structure of a separate message. If to use the system of relations of the native language, which we perceive automatically, then our head will not ache. However, for the linguistic processor to automatically identify them, one, probably, should work a bit.

### **PROSPECTS FOR THE FORMATION OF THE NATURAL LANGUAGE “DATA BASE” ON THE PRINCIPLES OF NEUROPHYSIOLOGY**

Then we will continue the study the process of forming the structural level of the “Data Base”, considering certain features of the organization of our neural network. The publication Kislenko–Sergeyev [9] presents a model of memory organization, where the cognitive element of the linguistic organization is the BSSS structure itself. There arises, of course, the question – how can a set of cognitive structures pass through a separate node of the neural network (analogue of the neuron)?

Here we come to the concept of “lexeme”, which, in practice, presents a model of a separate neuron, through which there can be many options for implementation of individual components of other BSSS. Lexema, practically defines all possible variants of usage of a wordform, which ensures the possibility of its entry to any level of the cognitive element of the linguistic message. For modern computer equipment, this is a simple procedure when we have a whole bank of lexemes for each language, as presented as the first step towards the formation of a cognitive structure.

The next step – there is “built” a corresponding framework of the BSSS structure (as an element of the cognitive web) based on the set of corresponding lexemes, through which many cognitive connections may pass. Thus, a real text is formed based on the separate of actant of lexemes, etc. (this is where the required lexemes are not in the dictionary). Continuing this procedure, we form a separate BSSS structure, through the “body” of which many connections may pass. That is, a separate lexeme can be a constructive element in creating a set of other similar structures. This is practically the procedure for the formation of a separate element of the BSSS structures, through which several other BSSS structures can pass. In general, such model represents integrational properties of the biological neuron.

To implement such functions, there was suggested to use the “marker” device (see Fig. 5); A separate BSSS structure, when reproducing a text, receives its unique marker (identifier). Thus, we have a model for forming a cognitive network of separate text, where each component of the message (except for BSSS) has its own unique number. Well, the procedure is complicated – but it can be performed automatically, freeing a person from this unskilled hard routine. This practically means that, with a limited number of lexemes (vocabulary), we will be able to reproduce virtually unlimited number of real texts. In a certain way, we are trying to implement the model of information accumulation in our neural network at the neural level. This is a virtually definite model of the neural network, where on the plurality of models of individual “neurons – lexemes” can be actualized a practically significant set of individual quants of knowledge in the form of BSSS structures, reproducing the cognitive potential of individual fields of knowledge – astronomy, mathematics, physics, etc.

Is it possible to implement such a cognitive neural network? - Basically – Yes. When taking into account the possibilities of forming and using powerful

distributed systems for the following functions: preservation of dictionaries of different languages, preservation of appropriate vocabularies of lexemes, formation of linguistic processors for the decomposition of arbitrary messages based on the BSSS structures, automatic marker formation system for each BSSS structure, a powerful translation system, where the elements of the translation are not words – but separate situation ... and so on.

What do we have in the final version? There is a large list of components for normal formation and operation of a whole cluster of Information Natural and Linguistic Technologies (INLT); a clear list of important tasks, each of which is being already solved as of date; there is only a lack in systematicity the entire chain of formation, processing and use of natural and linguistic information is not carefully researched.

Nevertheless, it is positive that these issues have already been put on the agenda, certain fields of research and design have been carefully described and investigated; young, ambitious, unbounded people join the development. It is also important that the set of lexemes of a particular area of research may be synthesized for a significant number of other quanta of knowledge, which provokes an ambitious question regarding the prospects of the possibility of automatic synthesis of new Knowledge. It is difficult, but this is probably the principle of the global artificial neural network, which is developed by the achievements in many scientific fields.

This model covers a plurality of models of individual quanta of knowledge presented at the level of the set of BSSS structures, which removes many problems, since a single quantum of knowledge is equally presented in all languages. At the same time, almost, all problems with implementation of such network are removed, if the basis of its formation is determined by the real text without any restrictions. Apparently, distributed information systems will be the basis for solving the problem of “Semantic Web – Data Base”. We can present a fragment of the neural organization of the proposed knowledge base structure (Fig. 6), where a separate element of the BSSS structure is updated with an appropriate lexeme associated with other lexemes of the reproduced text by a plurality of unique markers.

The linguistic processor, which allocates separate BSSS structures (for mono or poly- predicative organizations), translates them to the lexeme level, automatically forming a unique marker for each situation. This is, basically, a model of what is being implemented in our neural network on a plurality of neurons; this is a model of a neuron through which many relations with other quanta of knowledge can pass.

The main conclusion is considered rather optimistic: “there is proposed a model of automatic formation of the cognitive potential of the community in the form of a system of integration of all accumulated text material in various fields of human activity (physics, mathematics, biology, etc.) based on the standard BSSS structure as a unified scheme of textual information presentation”.

This means that we proceed to the constructive level of formation of the “cognitive potential of the society” in various areas of human intellectual activity, because we have structural capability of formation, accumulation and correction of knowledge presented at the linguistic level by a plurality of reports, research results, monographs, textbooks etc. Perhaps, the first research in this field is the

work of Toby Segaran “Programming the collective mind”, Symbol - Plus 368 p. [18] (unfortunately, I have not yet been able to get acquainted with this work). This is an interesting perspective direction of research, formation and generalization of collective achievement of humanity in different fields of activity.

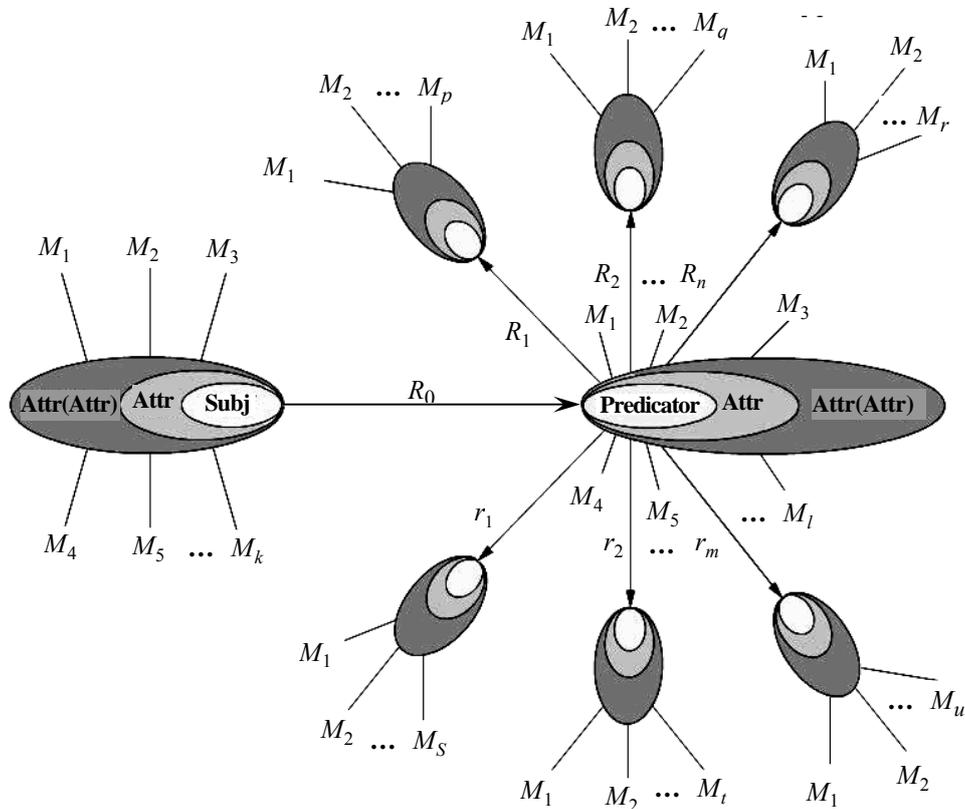


Fig. 6. The structure of the neural network fragment: Subj — BSSS subject; Attr (Subj) — subject’s attribute; Attr(Attr) — extend of the attribute; Predicator — core of BSSS structure; Attr(Pred) — predicative attribute; Attr(Attr(Pred)) — extend of the attribute of predicative

If the quantum of knowledge for the perception of a person of the environment presents a separate situation of a sensory level, and its verbalized form is the BSSS structure, then it will be a “cognitive element” of perception, accumulation and formation of the cognitive potential of the community, formed in separate areas of scientific research.

In practice, this section is devoted to the problems of formation of the “Data Base” on the principles of organization and operation of our neural network. In many ways, the proposed model functionally resembles the work of our neurosubstantium: perceives, “understands” and accumulates linguistic (text) information.

### SEARCH FOR THE SEMANTIC UNIT OF LINGUISTIC ACTIVITY

Basically, the previous section gave us an answer to this question. However, it is important to analyze the way of solving this issue. The presumption of formulation of such a question is determined by the rapid (powerful) development of computer technologies and the use of their capabilities, even for the simulation

of the most complex intellectual functions of a person. However, before transferring individual functions of intellectual activity to a computer, it is necessary: to clearly understand the statement of the problem and the question, the main principles of the linguistic organization, the scheme of language communication, reproduction of the environment by the means of language, "understand the message", etc., formalize these solutions for implementation at the computer level, or distribution systems.

1. The first stage has almost been performed by Sir Timothy Berners-Lee – he clearly set the problem of studying the linguistic activity as an acronym "SEMANTIC WEB – DATA BASE", which affects the main principles of the linguistic organization.

2. The second stage, initiated by Sir Timothy Berners-Lee, in the form of computer realization of this problem by using simplified speech message presentation models only through a plurality of triples (Subj, Pred, Obj); O.O. Varlamov has complicated the message presentation scheme by simultaneous use of the triples (Subj, Pred, Obj) and the logical functions "if ..  $S_1$  .. then ..  $S_n$ ", which connect separate situations in the form of messages of a monopredicative level.

3. The third stage (we can assume) is presented by the author's work ("Back to basics of speech activity") [8], it forms a new platform for studying the language activity, by introducing the concept of the Basic Semantic and Syntactic Structure (BSSS), which is derived from the structural and functional level of neural organization of the visual pathway and which will be the same for all humanity. That is, for the unit of perception, processing and "understanding" of language material it is necessary to take a separate BSSS, which simultaneously becomes a semantic unit of formation and presentation of an arbitrary text. The BSSS structure was presented at Fig. 5.

Of course, the question arises as to how complex is this presentation of natural language for a computer comparing with previous versions - the arguments here are as follows:

- 1) BSSS is a derivative from the structural and functional level of the neural organization of the visual pathway - hence, it will be the same for all languages;
- 2) BSSS is clearly and unambiguously formalized at the structural level;
- 3) it is completely conscious and correctly perceived by a person;
- 4) it appears as a structural unit for formation of an arbitrary message;
- 5) if the BSSS structure can be clearly and unambiguously presented at the linguistic level, then, apparently, in the same formalized way it can be translated to the computer level;
- 6) perhaps the main thing: it is possible to form a computer technology for processing PM texts in an automatic mode without a compulsory scheme for the formation of metatext for the computer.

## **FINAL PART**

Only in the process of working with the given topic (in fact, during the last final stage), the author himself realized the clear semantic load of the correlation between notions "Semantic Web – Data Base".

"Semantic Web" is a semantic matrix (standard relationship system), which we constantly use at the stages of analysis or synthesis of speech messages. This

is the level of our linguistic competence regarding the structural and functional levels of linguistic organization that provide us with and guarantee adequate character of implementation of the communication process in the modes of synthesis / analysis of speech communication. Our linguistic competence is formed, as they say “with mother’s milk” and the person / child masters the linguistic system at about 2,5-3 years, and confidently uses it at about 10-12 years.

There has been created (synthesized) nothing better than the “matrix of linguistic communication in the form of basic semantic and syntactic structure” by the nature, since this scheme is derived from the structural and functional level of the neural organization of the visual pathway; so, it is standard – the same for all humanity. This is a generalized scheme of how a person perceives all available information on the sensory or linguistic level, and the ability to translate it into the language level in the form of a messages of the mono/poly- predicative level.

Data Base, in this case, is a generalized scheme of accumulation, storage and presentation of the accumulated product by our linguistic activity for perpetual storage. And all attempt to change, transform, improve this communication scheme are useless. The only possible scheme of reproduction of the human cognitive potential is the BSSS structure. We should work for the computer to be able to understand this structural organization in the form of BSSS sequence.

Data Base is the big cognitive potential accumulated by humanity throughout the development of civilization, stored in papyrus, manuscripts, books, audio / video records. That is, all efforts should be directed for the computers, networks, distributed systems to store and generalize, but not lose the main values of our civilization.

Well, after carefully analyzing the structural level of the linguistic organization, we can come up with a more reasoned analysis of the individual language system (see Fig. 1), the components of which are the interconnected powers of the linguistic processor (LP) and the basic knowledge (BK). The first component defines the structural and functional levels of organization of the language channel as the basic scheme of perception and accumulation of the cognitive potential of humanity at the linguistic level, while the second – the BK defines the collective cognitive potential of a person, which stores all information both at the figurative level (obtained through the entire sensor system), and at the language level received through the perception of information by visual or audio channels. The latter, upon the request of the recipient, can always be translated to the language level.

The present study considers the functioning of LP and BK exclusively at the linguistic level. It is appropriate to emphasize the significant assistance of post-graduate students D.S. Serheiev and A. S. Khimicha in developing these issues. This is basically a team of like-minded people who have fruitfully worked on this ambitious program – modeling the speech behavior of a person in the form of an interconnected LP-BK system. The author expresses his sincere gratitude for cooperation in developing this area of research.

If Semantic Web is responsible for the process of perceiving and “understanding” the text, which, according to the integration approach, is based on a plurality of standard structures, then we can assume that the cognitive element of storage, and hence – the understanding and using the accumulated knowledge,

will be the basic semantic and syntactic structure itself. We should also note that the whole cognitive potential of a person is formed from a plurality of situations not only of symbolic, but also of figurative levels, preserving mother’s voice (with all its shades), appearance, character, all her attributes at the neural level.

The given work reproduces the results of the infinite stages of work on the declared “eternal” topic, and I should note the actual, or virtual attention and support that I constantly felt due to certain individuals.

In this way, I want to note the very pleasant audience and the spirit of communication at the “Artificial Intelligence” seminars, initiated by A.I. Shevchenko, the main priorities of which were: freedom of thought, freedom of imagination, goodwill and support for beginners. The last significant work of the author, “Personalized cognitive feedback, as a powerful lever of progressive acceleration of the development of society” was published due to support of Anatolii Ivanovych in the journal “Artificial Intelligence”, 2018, No. 1 [10].

Meetings and communication with the member of the academy O.O. Kryshtal were very productive. I would like to thank for acquaintance with his interesting treatises, covering the philosophical and linguistic problems of today, for his support to my search on the neurophysiological principles of linguistic activity; for submitting the article “Neurophysiological grounds of the structural organization of linguistic material”, which appeared at reports of the National Academy of Sciences of Ukraine, Edition 11, 2007 [6], and which further inspired the author for the monograph “From thought to knowledge” (neurophysiological grounds), publishing house “Ukrainian Chronicle”, 2008 [7].

It is important to emphasize the importance, urgency and timeliness of creation of the BICA (Biologically Inspired Cognitive Architectures) magazine, which is focused on studying the cognitive processes occurring in our neural network. For the first time we met (as I recall) with O.V. Samsonovich in Katsiveli at the seminar “Artificial Intelligence”; I think that from that day, he had an idea about creating the BICA magazine.

I am very grateful for the support and presentation of my research in this magazine: presentation of the research direction (BICA-13, Kyiv, Grandhotel), publication of the work “To the origins of speech activity” (BICA-14) [8], support of this research from my graduate students Yuri Kyslenko, Danylo Sergeiev – “Cognitive architecture of speech activity and modeling thereof” (BICA-15) [9].

In this context, one should note that when the main vector of social development (according to B.F. Porshnev) is a “progressive acceleration process”, we can predict that “the increase in the authority and the ranking of research of this direction will also occur under the scheme with a progressive acceleration”.

The last work, “Personalized cognitive feedback, as a powerful lever of progressive acceleration of the development of society” [10], confirms that only a living neurosubstance of a single individual is capable of synthesizing new information that was absent in the database. Whether it possible to perform it by IT technology in future, is a rhetorical question. But even this is very useful for improving the cognitive potential of the society. All knowledge that can be synthesized by living neurosubstances of the society and translated into the language level, may potentially become an achievement of the artificial intelligence, preserved and transmitted for future generations.

In this context, one should mention the words of the famous Kyiv citizen concerning the influence of the individual on the “progressive acceleration of the development of society”, written on the walls of his Alma Mater – the Igor Sikorsky Kyiv Polytechnic Institute: “

The work of an individual remains the spark that drives the society even more than collective work” (Igor Sikorsky).

## LITERATURE

1. *Anderson Janna*. The Fate of the Semantic Web / Janna Anderson, Lee Raine. — Pew Research Center’s Internet & American Life Project. — May 4, 2010.
2. *Щерба Л.В.* О трояком аспекте языковых явлений и эксперименте в языкознании. Языковая система и речевая деятельность / Л.В. Щерба. — М., 1974.
3. *Кисленко Ю.І.* Архітектура мови (лінгвістичне забезпечення інтелектуальних інтегрованих систем): навч. посібник / Ю.І. Кисленко. — К.: Віпол, 1998. — 343 с.
4. *Зеки Семір.* Зрительный образ в сознании и мозге: сборник трудов / Семір Зеки // В мире науки. — № 11–12. — М.: Мир, 1992. — С. 33–41.
5. *Хокинс Дж.* Об интеллекте / Дж. Хокинс, С. Блейкли. — М.: Издат. дом «Вильямс», 2007. — 240 с.
6. *Кисленко Ю.І.* Нейрофізіологічне підґрунтя структурної організації мовного матеріалу (Представлено академіком НАН України О.О. Кришталь) / Ю.І. Кисленко // Доповіді НАНУ. — 2007. — №11. — С. 158–164.
7. *Кисленко Ю.* От мысли к знанию (нейрофизиологические основания): моногр. / Ю. Кисленко. — К.: Издательство «Український літопис», 2008. — 102 с.
8. *Kislenko Y.I.* Back to basics of speech activity / Y.I. Kislenko // Biologically Inspired Cognitive Architectures. — 2014. — Vol. 8. — P.46–68.
9. *Kyslenko Yuri.* Cognitive architecture of speech activity and modeling thereof / Y.I. Kislenko // Biologically Inspired Cognitive Architectures. — 2015. — Vol. 12. — P.134–143.
10. *Kislenko Y.I.* Personified cognitive feedback as a powerful instrument for progressive acceleration of social evolution / Y.I. Kislenko // Штучний інтелект. — 2018. — № 1. — P.63–95.
11. *Port-Royal Grammar.* — 1966.
12. *Ф. де Соссюр.* Курс общей лингвистики. — М., 1963.
13. *Пиотровский Р.Г.* Лингвистические уроки машинного перевода / Р.Г. Пиотровский // Вопросы языкознания. — 1985. — № 4.
14. *Астахова Л.И.* Предложение и его членение: прагматика, семантика, синтаксис / Астахова Л.И. — Днепропетровский ГУ, 1992.
15. *Грамматика* современного русского литературного языка. — М.: Наука, 1970.
16. *Городецкий Б.Ю.* Компьютерная лингвистика: моделирование языкового общения / Пер. с англ. / сост., ред. и вступ. ст. Б.Ю. Городецкого // Серия «Новое в зарубежной лингвистике». — Вып 24. — М.: Прогресс, 1989. — 432 с.
17. *Гвоздев А.Н.* Формирование у ребенка грамматического строя русского языка / А.Н. Гвоздев. — М.: Изд-во АПН, 1949.
18. *Кисленко Ю.І.* Структурно-функціональний рівень організації лінгвістичного процесора / Ю.І. Кисленко, А.В. Хіміч // Системні дослідження та інформаційні технології. — 2018. — № 1. — С. 19–35.
19. *Ибука Масару.* После трех уже поздно / Масару Ибука. — М.: Знание, 2000. — 192 с.
20. *Кисленко Ю.І.* Проблемы и перспективы развития поисковых систем / Ю.І. Кисленко, А.В. Терентьев // Искусственный интеллект. — 2011. — № 3.
21. *Сегаран Тоби.* Программируем коллективный разум / Тоби Сегаран; пер. с англ. — Символ-Плюс. — 368 с.

Received 24.03.2019

---

From the Editorial Board: the article corresponds completely to submitted manuscript.