Acta Cybernetica **00** (0000) 1–21.

Constructive Model of the Natural Language

Viktor Shynkarenko[a], Olena Kuropiatnyk[a]

Abstract

The paper deals with the natural lenguage model. Elements of the model (the language constructions) are images with such attributes as sounds, letters, morphemes, words and other lexical and syntactic components of the language. Based on the analysis of processes of the world perception, visual and associative thinking, the operations of formation and transformation of images are pointed out. The model can be applied in the semantic NLP.

Keywords: image, image operations, constructive-synthesizing structure, natural language, language construct

1 Introduction

Information is one of the most important resources of the last decades. Considerable part of it is presented in verbal form in the natural languages (NL) and requires systematizing and automated processing to enable further acquisition of knowledge with the possibility of quick access to such knowledge. Systematization and further automation require formalization of the language concept and its components.

The problem of the natural language processing is conditioned by its features, such as permanent development including growth of the vocabulary and rules, redundancy, polysemy, and diversity of the forms of presentation.

There are many directions in the processing of texts in the NL (NLP) [8]: static and corps methods of NLP, usage of linguistic bases [22], [18], [33], finite state machines [19], [1] which is actively used, for example, in Nooj components [31], [32], regular expressions (in particular, in Semantic Tagger ANNIE Gate), and hidden Markov models [7].

Language as a set of constructions is represented: in the form of a neural probability model [2]; in the form of a tree-bank [17]; with the use of formulas of functional languages [11]; with the use of n-grams including those based on classes [5], and vector representation [14]. The latter approach is also used to create semantic models [15]. A graph presentation of semantics [22], [10], including semantic networks

[a]Department of Computer Informational Technologies, Dnipropetrovsk National University of Railway Transport named after Academician V. Lazaryan, Dnipro, Ukraine, E-mail: shinkarenko_vi@ua.fm, olena.kuropiatnyk@gmail.com

DOI: 10.14232/actacyb.00.0.0000.?

[6], thesauri based systems, tensor models [16], is also used for processing the texts in NL.

The paper is aimed to formalize the concept of language using the means of constructive-synthesizing structures [26]. The language will be represented in the form of the certain construction (in the broadest sense) which is the result of the corresponding constructive process of formation of images and words (phrases, sentences) as the attributes of images. As a result of this process, a constructive model of the NL is formed.

The paper represents development of the direction of mathematical and algorithmic constructivism which has already found its application in modelling the processes of alternatives ranking by the AHP method [24], adaptation of data structures in the operative memory [29], construction of a graph model of the text [28], etc.

2 State of the art

The main objective of NLP is the improvement of artificial intelligence (AI) systems. Today there are many sub-objectives aimed at improving the user interface with technical systems, quality of texts in the natural language, interaction between people, increasing the effectiveness of search engines and anti-plagiarism systems, etc.

Achieving the objectives involves solving of a number of problems: construction of models for representing the language and its constructions; development of methods for their processing and analysis, including specific applications (analysis of social media profiles for advisory systems, analysis of text messages, automatic translation and annotating...).

In the course of solving these tasks and achieving the objectives, the following questions arise: NL coverage (vocabulary, syntax, and semantics); possibilities of expanding the language model; ability to work with polysemous words, synonyms, homonyms (homographs); possibility of oral speech modelling (including recognition of homophones), taking into account personality of a speaker, approaching the natural thinking processes of an individual.

Models based on n-grams and tree-banks cover vocabulary and syntax and find their usage in Stanford Parser [21]; semantics is represented in graph and vector models [22], [15], [10]. All the models considered allow scaling the model. Modification of n-gram models due to probabilities [2] reduces their dimensionality. The possibility of modeling non-written speech is assumed in the models proposed in the works of Krak [12], [13]. There are no models used in NLP which take into account peculiarities of an information source. Models allowing classification of the language [9] also work regardless of its carrier.

A model that is close to the processes of human mental activity is considered in the work [4]; it is based on the figurative analysis and synthesis. Also, there is an approach proposed for constructing a conceptual model of the figurative analysis and synthesis of NL structures on the basis of psycho-physiological phenomena [3].

Studies have shown that NLP models work with texts, without taking into account specific features of their authors. At the same time, they are aimed at studying the concepts of the language and text as its construction. They are the most complete and close in its semantic nature to the studies carried out by the authors.

The proposed language model is based on the model of the human image system, and operations introduced for its construction and expansion do not contradict the above mentioned ones. Therefore, we have an opportunity to cover both the vocabulary and semantics of the oral and written speech, as well as to take into account phonetic and personal characteristics of the language and the particular individual, respectively.

3 Generalized constructive-synthesizing structure

The following triple [26] is called the generalized constructive-synthesizing structure (GCSSt):

$$C_G = \langle M,\ \Sigma,\ \Lambda \rangle\,,$$

where M is inhomogeneous structure medium (the main set of the elements), Σ is the signature comprising a set of relations and linking, substitution and output operations, as well as operations on attributes, Λ is the constructive axiomatics. GCSSt axiomatics is presented in the paper [26].

The constructive-synthesizing structure (CSSt) is intended for the formation of a plurality of structures using operations and relations of signatures, the rules for implementation of which are given in the axiomatics.

To form the structures, it is necessary to perform a number of transformations of CSSt: specialization, interpretation, and concretization [27]. Implementation of CSSt consists in formation of constructions (in this context, language constructs) of elements of CSSt medium by performing CSSt algorithms related to operations of the signature.

4 Specialization of CSSt of human images

Everything that surrounds an individual, the real and the virtual things (processes, entities, events and phenomena), as well as the individual him/herself, including material (tissues, organs) and nonmaterial (emotion, feelings) components, will be called the prototype, as a certain integral part of the world which is considered in isolation.

Image sensitivity is a characteristic feature of any individual. We understand the image as a representation of the prototype, its properties on some physical medium. Such a medium can be an individuals memory as a part of the nervous system, animals memory, computer, and computer networks.

Specialization involves determining the application environment, i.e., semantic nature of the CSSt medium, a finite set of operations and their semantics, operation attributes, as well as the order of their performing. Let us consider CSSt specialization of the human image system:

$$C = \langle M,\, \Sigma,\, \Lambda \rangle_{\,S} \mapsto {}_{S}C_h = \langle M_h,\, \Sigma_h,\, \Lambda_h \rangle\,, \tag{1}$$

where $M_h = T \bigcup N$ is a heterogeneous scalable medium, T is a set of terminals – images, N is a set of non-terminals, Σ_h is the signature of relations and operations performed on elements of the medium, Λ_h is the constructive axiomatics containing updates, additions and restrictions for media elements, operations and signature relations, on the basis of which the construction is performed.

4.1 Partial axiomatics of the medium

The image ${}_{\bar{w}}m_i \in M_h$ has a set of attributes $\bar{w} = \{w_1, w_2, \ldots, w_n\}$. Heterogeneous multiset of elements with attributes is meant by the set. Belonging of the attribute w_j to the image m will be denoted as $w_j \;\underset{\leftarrow}{\rfloor}\; m$. All attributes are the images.

The images may change over time. Each image has an attribute of the time of creation or last modification ($t \;\underset{\leftarrow}{\rfloor}\; m_i$). The given attribute is changed in the course of operation on the image and depends on the time of its execution.

The world image ${}_{t}P \in M_h$ will be called a continuous representation of the human environment presented in the form of dynamic flow of images, sounds, tactile, gustatory, olfactory and spatial-temporal sensations, feelings and emotions, reflected by the nervous system of an individual under the influence of physical stimuli. This image is a controlled one and depends on any particular individual (it is not essential in the context of this paper). At any given time t the certain world representation exists. Further, this attribute is not specified.

The form (l) is a set of elements M_h connected by the relationships of Σ_h.

Sentential form is a form obtained at any time as a result of inference from the initial non-terminal symbol according to the rules of inference from concretized CSSt.

The construction (K) is a sentential form at the current time, comprising only the terminals [26]. Constructions and relations are images as well.

The set of images is a construction based on some relation of similarity with the properties of reflexivity and symmetry [23].

4.2 Partial axiomatic of operations and relations

The signature Σ_h consists of the set of operations $\Sigma_h = \langle \Xi,\, \Theta,\, \Phi,\, \{\rightarrow\}\rangle \bigcup \Psi$, where Ξ – relations and homonym operations, operations of linking and transformation of the medium elements $\{\cdot,\, \vec{\in},\, \circ,\, \bar{\circ},\, \wedge,\, \vee,\, \diamondsuit, \uparrow,\, \Uparrow,\, >>,\, <<,\, \exists\} \subset \Xi$, $\Theta = \{\Rightarrow,$ $|\Rightarrow,\, ||\Rightarrow\}$ – the operations of substitution and inference, $\Phi = \{:=\}$ – the relations and homonym operations on the attributes, $\{\rightarrow\}$ – substitutive relation. $\Psi = \{\psi_i : \langle s_i,\, g_i\rangle\}$ is a set of substitution rules, s_i – a sequence of substitutive relations, g_i – set of operations on the attributes.If the operations on attributes are not performed, the substitution rule will take the form $\langle s_i,\, \varepsilon\rangle$, where ε is a null character. Relations from Ξ are applied in the inference rules, and operations corresponding to relations are applied during implementation of CSSt.

Execution time is an essential attribute of any operation $\tau \hookleftarrow *$, where $*$ means any operation of Σ_h. Time attribute of each image after operation can be determined as $t = t_{start} + \tau$, with t_{start} representing the time of the operation start. The value of this attribute is determined by abilities of the performer. Further, it is not specified.

Image concatenation operation $\cdot({}_{\bar{w}_1}m_1,\ {}_{\bar{w}_2}m_2,\ P)$ involves linking of images ${}_{\bar{w}_1}m_1,\ {}_{\bar{w}_2}m_2$ under the influence of the world image P. The result is an image m – a sequence of the images ${}_{\bar{w}_1}m_1,\ {}_{\bar{w}_2}m_2$.

Element inclusion operation $\vec{\in}(\bar{m},\ {}_{\bar{w}}m)$ involves adding the image ${}_{\bar{w}}m$ into a set $\bar{m}$, $\bar{m}$ is the operand and result of the operation.

Image explication operation [25] is a selection of the part of the world and formation of the individual object-image with its own set of attributes. Result of the operation $\circ({}_{\bar{w}_1}m_1,\ {}_{\bar{w}_2}m_2)$ is the image ${}_{\bar{w}}m$ explicated from the image ${}_{\bar{w}_1}m_1$ under the influence of ${}_{\bar{w}_2}m_2$. The images ${}_{\bar{w}_1}m_1,\ {}_{\bar{w}_2}m_2$ can be any images of the medium M_h, construction or image of the world ${}_{focus,\,t}P$ at some moment in time t, on which attention of an individual is focused (it is indicated by the attribute $focus$). Further these attributes of the image P will be used as needed. Modification of the operation $\bar{\circ}({}_{\bar{w}_1}m_1,\ {}_{\bar{w}_2}m_2)$ is an explication of the relation ${}_{\bar{w}}m$ from ${}_{w}m_1$ under the influence of the image ${}_{\bar{w}_2}m_2$.

Inheritance operation with specification $\wedge({}_{w_1}m_1,\ {}_{w_2}m_2)$ involves the creation of a new image ${}_{w^*}m^*$ that repeats the image ${}_{\bar{w}_1}m_1$ and has $\bar{w}_1$ and ${}_{\bar{w}_2}m_2$ as the attributes.

Inheritance operation with the modification $\vee({}_{\bar{w}_1}m_1,\ {}_{\bar{w}_2}m_2 \hookleftarrow m_1,\ {}_{\bar{w}_3}m_3)$ involves creation of the new image ${}_{w^*}m^*$ that repeats the image ${}_{\bar{w}_1}m_1$ with substitution of the attribute ${}_{\bar{w}_2}m_2$ for ${}_{\bar{w}_3}m_3$.

The finite set of linking operations of the images ${}_{\bar{w}_i}m_i,\ {}_{\bar{w}_j}m_j - \Diamond_k({}_{\bar{w}_i}m_i,\ {}_{\bar{w}_j}m_j)$. The result of the operation is m', composite image or relation image. Each of the given operations is the image of relation and belongs to the medium and signature $(\Diamond_i \in M_h,\ \Diamond_i \in \Sigma_h)$.

Generalization operation $\uparrow(\bar{m})$, where $\bar{m}$ is a set of images involves selection of the set with one or more identical attributes and formation of the result of operation m as a new generalized image with the same attributes and similar transformed images of the original images.

Unification operation ${}_c \Uparrow (\bar{m},\ {}_{\bar{w}_1}m_1,\ {}_{\bar{w}_2}m_2)$ allow creating some set of images $\bar{m}$ with adding of the image ${}_{\bar{w}_1}m_1$, provided that each element $\bar{m}$ and image ${}_{\bar{w}_1}m_1$ have similar attribute ${}_{\bar{w}_2}m_2$.

Image transfer operation ${}_{ch} >> ({}_{\bar{w}}m,\ P)$ involves transmission of the image ${}_{\bar{w}}m$ to the external world through the channel ch in some encoded form. In the course of forming the language, it is visual (s) and auditory (h) form. Upon that, the world image P is changed, being supplemented by a new image of the operation performer image of the code: word, sentence, gesture, etc.

Image reception operation ${}_{ch} << (P^*)$ involves determining (obtaining) certain code of the image ${}_{\bar{w}}m$ using the channel ch from the part of the external world P^*.

In this image of the world P is not changed, and the medium of CSSt images of performer is supplemented by a new image.

Operation to verify existence of the attribute $\exists({}_{\bar{w}_1}m_1,\ {}_{\bar{w}_2}m_2)$ determines presence of the attribute ${}_{\bar{w}_2}m_2$ in the image ${}_{\bar{w}_1}m_1$, i.e. ${}_{\bar{w}_2}m_2 \hookleftarrow {}_{\bar{w}_1}m_1$. The result of the operation is the logical value of truth in the presence of the required attribute ${}_{\bar{w}_2}m_2$, otherwise it is false.

Substitutive relation is a binary relation with the attributes ${}_{w_i}l_{iw} \to {}_{w_j}l_j$, where l_i, l_j are the sentential forms [26]. The sequence of the substitutive relations s_n is written as $s_m = \langle l_i \to l_j | l_k \rangle$, where l_i, l_j, l_k are the sentential forms. The substitutive relation can be written in short form $s_m = \langle l_i \to l_j | l_k \rangle$, where l_i, l_j, l_k are the forms, and it is equivalent to $s_m = \langle l_i \to l_j \rangle$, $s_n = \langle l_i \to l_k \rangle$.

For the given form ${}_{w_l}l = {}_{w_0} \otimes ({}_{w_1}l_1,\ {}_{w_2}l_2,\ \ldots,\ {}_{w_h}l_h,\ \ldots,\ {}_{w_k}l_k)$ and available substitutive relation ${}_{w_p} \to ({}_{w_h}l_h,\ {}_{w_q}l_q)$, where ${}_{w_h}l_h$ is a subform ${}_{w_l}l$ (${}_{w_h}l_h \prec {}_{w_l}l$), the result of the triple substitution operation ${}_{w_l^*}l^* = {}_{w_p} \Rightarrow ({}_{w_h}l_h, {}_{w_q}l_q,\ {}_{w_l}l)$ will be the form ${}_{w_l^*}l^* = {}_{w_0} \otimes ({}_{w_1}l_1,\ {}_{w_2}l_2,\ \ldots,\ {}_{w_q}l_q,\ \ldots,\ {}_{w_k}l_k)$, where $\Rightarrow \in \Theta$, $\otimes$ is any operation of linking from Ξ.

Double operation of partial output ${}_{w_l^*}l^* = {}_{w_p}| \Rightarrow (\Psi,\ {}_{w_l}l)$ ($| \Rightarrow \in \Theta$) consists in:

- selection of one of the available substitution rules $\psi_r\ :\ \langle s_r,\ g_r \rangle$ with the substitutive relations s_r;
- performance of substitution operations on the basis of it;
- performance of on the attributes g_r in the predetermined sequence.

Binary operation of full output or simply output ($|| \Rightarrow (\Psi,\ {}_{w_l}l)$, $|| \Rightarrow \in \Theta$) resents step-by-step transformation of forms, starting from the initial non-terminal and ending with the construction satisfying the condition of the output ending, which implies cyclical performance of the partial output operations.

Operation $:= (a,\ b)$ consists in assigning the value of operand b to the operand a.

5 Interpretation of CSSt of human images

To determine the performance algorithms of possible operations and relations on images, let us interpret the structure (1):

$$\begin{aligned}\langle C_h = \langle M_h,\, \Sigma_h,\, \Lambda_h \rangle\,,\, C_A = \langle M_A,\ \Sigma_A,\ \Lambda_A \rangle \rangle_{\,I} \mapsto {}_{\,I,\,C_A} \\ \mapsto {}_{\,I,\,C_A}C_h = \langle M_h,\, \Sigma_h,\, \Lambda_1,\, Z \rangle\,,\end{aligned} \tag{2}$$

where $M_A \supset V_A$, $V_A = \{A_i^0|_{X_i}^{Y_i}\}$ – a set of basic algorithms [30], X_i, Y_i – sets of determinations and values of the algorithm $A_i^0|_{X_i}^{Y_i}$, $\Lambda_1 = \Lambda_h \bigcup \Lambda_A \bigcup \Lambda_2$, Z – a set of possible CSSt performers which are able to implement all algorithms C_A; $\Lambda_A = \{M_A = \bigcup\limits_{A_i^o \in V_A} (X(A_i^0) \bigcup Y(A_i^0)) \bigcup \Omega(C_h)\}$ – inhomogeneous medium, $\Omega(C_h)$ – a set of constructions of the images which satisfy C_h.

Performer ${}_{\bar{k}_i}z_i$ of the structure (2) has a set of attributes, and we shall distinguish some of them $\bar{k} = \{location,\ occupation,\ l\_condition,\ p\_characters\}$, where *location* means the locality (residence), *occupation* means profession (activity), *l_condition* means the living conditions, *p_characters* are psycho-physiological characteristics, including those connected with perception and processing of information.

The structure ${}_{I,\,C_A}C_h$ includes algorithms of performing the operations:

- A_1^0 – composition of algorithms $A_1^0|_{A_i,\,A_j}^{A_i \cdot A_j}$, $A_i \cdot A_j$ – sequential execution of the algorithm A_j after algorithm A_i;
- A_2^0 – conditional execution $A_2^0\,|_{b}^{A_1}$: algorithm A_i is performed, if the condition b is true (execution is allowed);
- $A_3|_{m_1,\,m_2,\,P}^{m_1 \cdot m_2}$ – images concatenation;
- $A_4|_{m,\,\bar{m}}^{\bar{m}}$ – image inclusion;
- $A_5|_{m_1,\,m_2}^{m^*}$ – image explication;
- $A_6|_{m_1,\,m_2}^{m^*}$ – explication of relation image;
- $A_7|_{m_1,\,m_2}^{m^*}$, $A_8|_{m_1,\,m_2,\,m_3}^{m^*}$ – inheritance of the image with specification and modification;
- $A_9|_{m_i,\,m_j}^{m'}$ – linking the images (establishment of relations);
- $A_{10}|_{\bar{m}}^{\bar{m}}$ – images generalization;
- $A_{11}|_{\bar{m},\,m}^{\bar{m}}$ – images unification;
- $A_{12}|_{m,\,P}^{P}$, $A_{13}|_{m,\,P}^{P}$ – transfer of image using audio and visual channel;
- $A_{14}|_{P}^{m}$ – receiving the image using audio and visual channel;
- $A_{16}|_{m_1,\,m_2}^{c}$ – checking for presence of the attribute m_2 in the image m_1;
- $A_{17}|_{l_h,\,l_q,\,f_i}^{f_i}$ – substitution;
- $A_{18}|_{f_i,\,\Psi}^{f_j}$, $A_{19}|_{\sigma,\,\Psi}^{\bar{\Omega}}$ – partial and full output, where f_i, f_j – forms, σ – axiom, $\bar{\Omega}$ – a set of the formed constructions;
- $A_{20}|_{a,\,b}^{b}$ – assignation.

Axiomatics of linking the operations and algorithms is as follows:
$\Lambda_2 = \{(A_1^0|_{A_i,A_j}^{A_i \cdot A_j} \hookleftarrow \cdot),\ (A_2^0\,|_{b}^{A_1} \hookleftarrow :), (A_3|_{m_1,\,m_2,\,P}^{m_1 \cdot m_2} \hookleftarrow \cdot), A_4|_{m,\,\bar{m}}^{\bar{m}} \hookleftarrow \vec{\in}),$
$(A_5|_{m_1,\,m_2}^{m^*} \hookleftarrow \circ), (A_6|_{m_1,\,m_2}^{m^*} \hookleftarrow \bar{\circ}), (A_7|_{m_1,\,m_2}^{m^*} \hookleftarrow \wedge), (A_8|_{m_1,\,m_2,\,m_3}^{m^*} \hookleftarrow \vee),$
$(A_9|_{m_i,\,m_j}^{m'} \hookleftarrow \Diamond), (A_{10}|_{\bar{m}}^{\bar{m}} \hookleftarrow \uparrow), (A_{11}|_{\bar{m},\,m}^{\bar{m}} \hookleftarrow \Uparrow), (A_{12}|_{m,\,P}^{P} \hookleftarrow {}_h >>),$

$(A_{13}|^{P}_{m,\,P} \hookleftarrow\!\rfloor_{\,s} >>), (A_{14}|^{m}_{P} \hookleftarrow\!\rfloor_{\,h} <<), (A_{15}|^{m}_{P} \hookleftarrow\!\rfloor_{\,s} <<), (A_{16}|^{c}_{m_1,\,m_2} \hookleftarrow\!\rfloor \exists),$
$(A_{17}|^{f_i}_{l_h,\,l_q,\,f_i} \hookleftarrow\!\rfloor \Rightarrow), (A_{18}|^{f_j}_{f_i,\,\Psi} \hookleftarrow\!\rfloor \mid \Rightarrow), (A_{19}|^{\bar{\Omega}}_{\sigma,\,\Psi} \hookleftarrow\!\rfloor \mid\mid \Rightarrow), (A_{20}|^{b}_{a,\,b} \hookleftarrow\!\rfloor :=)\}.$

These algorithms are specific to each internal performer; they are based on the chemical and biological processes associated with the work of the human nervous system (partially they are highlighted in the attributes). Since the features of these operations performance depend on the performer, the multiple interpretation supposing different algorithms for one and the same operation is possible.

6 Concretization of CSSt of the human images

To clarify the input operations, let us perform concretization of the structure (2):

$$_{I,\,C_A}C_h = \langle M_h,\, \Sigma_h,\, \Lambda_1,\, Z\rangle \;_K\!\mapsto_{K,I,\,C_A} C_h = \langle M_h,\, \Sigma_h,\, \Lambda_2,\, Z\rangle\,, \tag{3}$$

where $\Lambda_2 = \Lambda_1 \bigcup \Lambda_3$, $\Lambda_3 \supset \{M_h = T \bigcup N,\, T = \{K,\, P,\, K_{pw},\, K_s,\, K_{aw}\}\}$ – a set of terminals, K – construction in the form of the set of images, K_s – construction of the ordered images of sounds, K_{pw} – construction of the ordered images of the written construction, K_{aw} – construction of the images received during observation of actions, glances, facial expressions and so on, $N = \{\sigma,\, \eta,\, \alpha,\, \beta,\, \delta,\, \chi,\, \gamma,\, \kappa\,\mu,\, \theta,\, \nu,\, \lambda\}$ – a set of non-terminals, σ – initial non-terminal.

6.1 Axiomatics of substitution rules

Let us consider the operations associated with imaginative thinking of an individual.

Rules of substitution $s_1 - s_3$ allow generating a new image based on the explication operation:

$$s_1 = \langle \sigma \to K\rangle\,, s_2 = \left\langle K \to \vec{\in}(K,\chi)\right\rangle, s_3 = \langle \chi \to \circ(P,\,\varepsilon)| \circ (K,\,P)| \circ (P,\,K)\rangle\,.$$

Rules of substitution $s_4 - s_5$ allow performing inheritance of the image with the specification:

$$s_4 = \left\langle K \to \vec{\in}(K, \wedge(\chi,\,\gamma))\right\rangle, s_5 = \left\langle \gamma \to \circ(P,\,K)| \circ (K,\,P),\, K \to \vec{\in}(K,\gamma)\right\rangle.$$

Rules of substitution $s_5 - s_7$ allow inheriting the image with the modification

$$s_6 = \left\langle K \to \vec{\in}(K, \vee(\chi,\, \beta \hookleftarrow\!\rfloor\, \chi,\, \gamma)\right\rangle, s_7 = \langle \beta \to \circ(\chi,\,P)| \circ (\chi,\,K)\rangle\,.$$

Rules of substitution s_8, s_9 can be used for generalization of images:

$$s_8 = \left\langle K \to \vec{\in}(K, \uparrow(\alpha)\right\rangle, s_9 = \langle \alpha \to_c \Uparrow (\alpha,\,\gamma,\,\beta)|\varepsilon\rangle\,, g_9 = \langle c := \exists(\gamma,\,\beta)\rangle\,.$$

Image detailing is the operation being the reverse of generalization. It is performed by adding new attributes to generalized image using the specification operation. The following substitutive relations make it possible to link two images:

$$s_{10} = \left\langle K \to \vec{\in}(K,\,\mu),\, \mu \to \Diamond(\chi,\,\gamma),\, \Diamond \to \bar{\circ}(P,\,K)\right\rangle,$$

$$s_{11} = \left\langle K \to \vec{\in}(K,\, \mu),\, \mu \to \diamondsuit(\chi,\, \gamma),\, \diamondsuit \to \bar{\circ}(K,\, P)\right\rangle,$$

where $\diamondsuit \to \bar{\circ}(P,\, K)$, $\diamondsuit \to \bar{\circ}(K,\, P)$ is the selection of the relation image from the external world or construction, $\chi,\, \gamma \in M_h$.

Further we shall consider the operations related to the associative thinking and transmission of information.

Substitutive relations $s_{12} - s_{15}$ can be used to determine the attribute of the code (η) for the image χ:

$$s_{12} = \left\langle K \to \vec{\in}(K,\, \wedge(\chi,\, \eta)),\, \eta \to \wedge(\eta,\, \chi)\right\rangle, s_{13} = \langle \eta \to \circ(K_s,\, K) | \circ (K_s,\, P)\rangle\,,$$

$$s_{14} = \left\langle K \to \vec{\in}(K,\, K_s),\ K_s \to \vec{\in}(K_s,\, \nu)\right\rangle,$$

$$s_{15} = \Big\langle \nu \to \cdot(\delta,\, \nu,\, P),\, \delta \to \circ(sound \lrcorner P,\, P) | \circ (sound \lrcorner P,\, K)$$
$$| \circ (sound \lrcorner K,\, K) | \circ (sound \lrcorner K,\, P) | \varepsilon,\ \nu \to \circ(sound \lrcorner P,\, P)$$
$$| \circ (sound \lrcorner P,\, K) | \circ (sound \lrcorner K,\, K) | \circ (sound \lrcorner K,\, P)\Big\rangle\,.$$

The code can be used for transmission of information and in the process of thinking. The code may be represented as the image of sound (s_{15}) or picture (s_{19}) originally selected from the image of the external world. The sound image is a construction built on single atomic sounds, i.e. phonemes.

The visual image (the letter) can be put in correspondence with the image of the phoneme; the language construction (LC), such as word, word combination, sentence, etc., can be associated with the image of sound.

The written LC can be constructed as follows:

$$s_{16} = \left\langle K \to \vec{\in}(K,\, \wedge(\chi,\, \kappa)),\, \eta \to \wedge(\kappa,\, \chi),\, \kappa \to \wedge(\kappa,\, \eta), \eta \to \wedge(\eta,\, \kappa)\right\rangle,$$

$$s_{17} = \langle \kappa \to \circ(K_{pw},\, K) | \circ (K_{pw},\, P)\rangle\,,$$

$$s_{18} = \left\langle K \to \vec{\in}(K,\, K_{pw}),\, K_{pw} \to \vec{\in}(K_{pw},\, \kappa)\right\rangle,$$

$$s_{19} = \Big\langle \kappa \to \cdot(\delta,\, \kappa,\, P),\, \delta \to \circ(img \lrcorner P,\, P) | \circ (img \lrcorner P,\, K) | \circ (img \lrcorner K,\, K) |$$
$$\circ(img \lrcorner K,\, P) | \varepsilon,\, \kappa \to \circ(img \lrcorner P,\, P) | \circ (img \lrcorner P,\, K) | \circ (img \lrcorner K,\, K) |$$
$$\circ(img \lrcorner K,\, P)\Big\rangle\,,$$

where $img \lrcorner P$ are the pictures included in the image of the world.

As a result of implementation of the rules $s_{12} - s_{19}$, the image of the external world P is put in correspondence with each completed language and visual construction. The image of the external world is complemented by the LC attribute identifying the same.

In addition to speech and written LC the character constructions of images (K_{aw}), such as gestures, glances, facial expressions, special fonts and scripts (for example, Braille script) and other actions and sensations can be formed:

$$s_{20} = \left\langle K \to \vec{\in}(K,\, \theta),\, \theta \to \circ(K_{aw}, K) | \circ (K_{aw}, P)\right\rangle,$$

$$s_{21} = \left\langle K \to \vec{\in}(K,\, K_{aw}),\, K_{aw} \to \vec{\in}(K_{aw},\, \lambda) \right\rangle,$$

$$s_{22} = \Big\langle \lambda \to \cdot(\theta,\, \lambda,\, P),\, \theta \to \circ(imgd \,{}_{\hookleftarrow}\!\rfloor\, P, P)| \circ (imgd \,{}_{\hookleftarrow}\!\rfloor\, K, P)| \circ (imgd \,{}_{\hookleftarrow}\!\rfloor\, P, K)|$$
$$\circ(imgd \,{}_{\hookleftarrow}\!\rfloor\, K, K)|\varepsilon,\;\; \lambda \to \circ(imgd \,{}_{\hookleftarrow}\!\rfloor\, P, P)| \circ (imgd \,{}_{\hookleftarrow}\!\rfloor\, K, P)| \circ (imgd \,{}_{\hookleftarrow}\!\rfloor\, P, K)|$$
$$\circ(imgd \,{}_{\hookleftarrow}\!\rfloor\, K, K) \Big\rangle,$$

where *imgd* is dynamic image associated with some human activities.

The substitutive relation s_{23} allows transfer images:

$$s_{23} = \langle \chi \to {}_{ch} >> (\eta,\, P) \rangle\,.$$

Substitutive relations $s_{24} - s_{25}$ allow receiving the image by supplementing the CSSt medium of LC images of a performer (an individual) in the following forms:

- written form $s_{24} = \langle K \to \vec{\in}(K,\, K_{pw}),\, K_{pw} \to \vec{\in}(K_{pw},\, \kappa),$ $\kappa \to {}_{s} << (img \,{}_{\hookleftarrow}\!\rfloor\, P),\;\; K \to \vec{\in}(K,\, \wedge(\chi,\, \kappa)),\, \chi \to \circ(P,\, \varepsilon)| \circ (K,\, P)| \circ (P,\, K),$ $K_{pw} \to \vec{\in}(K_{pw},\, \wedge(\kappa,\, \chi))\rangle;$

- speech form $s_{25} = \langle K \to \vec{\in}(K,\, K_s),\, K_s \to \vec{\in}(K_s,\, \nu),$ $\nu \to {}_{h} << (sound \,{}_{\hookleftarrow}\!\rfloor\, P),\;\; K \to \vec{\in}(K,\, \wedge(\chi,\, \nu)),\;\; \chi \to \circ(P,\, \varepsilon)| \circ (K,\, P)| \circ$ $(P,\, K),\;\; K_s \to \vec{\in}(K_s,\, \wedge(\nu,\, \chi))\rangle;$

- other form $s_{26} = \langle K \to \vec{\in}(K,\, K_{aw}),\;\; K_{aw} \to \vec{\in}(K_{aw},\, \theta),\, \theta \to_s << (imgd \,{}_{\hookleftarrow}\!\rfloor$ $P),\, K \to \vec{\in}(K,\, \wedge(\chi,\, \theta)),\, \chi \to \circ(P,\, \varepsilon)| \circ (K,\, P)| \circ (P,\, K),$ $K_{aw} \to \vec{\in}(K_{aw},\, \wedge(\theta,\, \chi)).$

7 Application of CSSt for constructive modeling of images

Let us consider the example of receiving and constructing images for a specific language construct – sentence 1 – "The branch operator is an operator that ensures the performing of certain commands only if a certain logical expression is true".

This sentence is perceived as a construction of visual images-symbols consisting of images of words. To process this construction it is necessary to:

1. obtain all images of words using the operation ${}_{s} << (img \,{}_{\hookleftarrow}\!\rfloor\, P)$, where P is the image of the world, which includes the sentence under consideration;

2. compare the images of words with the images of prototypes named by them. Words-articles do not have any significant influence on the meaning of constructions, hence their semantic images will be omitted;

3. build a construction of images expressing the meaning of the language construct.

Let us implement the second paragraph:

$$K \overset{24(1)}{\Rightarrow} \vec{\in}(K,\, K_{pw}) \overset{24(2)}{\Rightarrow} \vec{\in}(K,\, \vec{\in}(K_{pw},\, \kappa_i)) \overset{24(3)}{\Rightarrow} \tag{4}$$

$$\overset{24(3)}{\Rightarrow} \vec{\in}(K, \vec{\in}(K_{pw},\, {}_s << (img \hookleftarrow\rfloor\, {}_{f_i,\, t_i}P) = K_i, i = \overline{1, C}$$

where κ_i is the image of the word added to the performer's images, K_i is the construction of images – result of fulfillment of the relations, C is the number of words in the sentence, 24(1) is the application of the first relations from the set of relations s_{24} (likewise for similar records).

Let us establish the connection "meaning-word":

$$K_i \overset{24(4)}{\Rightarrow} \vec{\in}(K_i,\, \wedge(\chi_i,\, \kappa_i)) \overset{24(5)}{\Rightarrow} \vec{\in}(K_i,\, \wedge(\circ(K_i,\, {}_{f_i,t_i}P)_i,\, \kappa_i)), \tag{5}$$

where κ_i is the image of the word that is an independent part of speech.

The connection "word-meaning" (rule 24(6)) will be established if the image of the word is received by performer for the first time. Conclusions similar to (4-5) can be further omitted.

Let us construct the images corresponding to the sentence under consideration, using the formula (4). We shall form the images of words $\kappa_1 - \kappa_{21}$. The time and focus corresponding to this operation are given below (Table 1).

Table 1: Images of words of the sentence 1.

Image / word	Focus	Time	Image /word	Focus	Time
1/the	1	1	11/of	14	14
2/branch	2	2	12/certain	15	15
3/operator	3	3	13/commands	16	16
4/is	5	5	14/if	21	21
5/an	6	6	15/a	22	2
6/operator	7	7	16/certain	23	23
7/that	9	9	17/logical	24	24
8/ensures	10	10	18/expression	25	25
9/the	11	11	19/is	28	28
10/the	12	12	20/true	29	29

For all words of speech, we explicate images-meanings χ_i, $i = \overline{1,\, 20}$ and connect them with the words using the formula (5). The time and focus will coincide with the corresponding indicators when receiving word images. When constructing the images (complex and composite images), images of non-independent parts of speech can be omitted or interpreted as images of relations. Let us consider construction of composite images (Table 2). The constructed images can be enriched by adding attributes and their specification.

Table 2: Constructing the composite image constructs.

Conclusion	Image	Focus	Time	Prototype
$K \overset{4}{\Rightarrow} \vec{\in}(K, \wedge(\chi_2, \chi_1))$		4	4	the branch operator
$K \overset{11(1)}{\Rightarrow} \vec{\in}(K, \mu_1)$ $\overset{11(2)}{\Rightarrow} \vec{\in}(K, \Diamond_1(\wedge(\chi_2, \chi_1), \chi_6))$ $\overset{11(3)}{\Rightarrow} \vec{\in}(K, \bar{\circ}(K, {}_{f_8, t_8}P)$ $(\wedge(\chi_2, \chi_1), \chi_6))$	μ_1	8	8	the branch operator is an operator
$K \overset{4}{\Rightarrow} \vec{\in}(K, \wedge(\chi_{13}, \chi_{12}))$		17	17	certain commands
$K \overset{11(1)}{\Rightarrow} \vec{\in}(K, \mu_2)$ $\overset{11(2)}{\Rightarrow} \vec{\in}(K, \Diamond_2(\chi_{10}, \wedge(\chi_{13}, \chi_{12})))$ $\overset{11(3)}{\Rightarrow} \vec{\in}(K, \bar{\circ}(K, {}_{f_{18}, t_{18}}P)$ $(\chi_{10}, \wedge(\chi_{13}, \chi_{12})))$	μ_2	18	18	the performing of certain commands
$K \overset{11(1)}{\Rightarrow} \vec{\in}(K, \mu_3)$ $\overset{11(2)}{\Rightarrow} \vec{\in}(K, \Diamond_3(\chi_8, \mu_2)) \overset{11(3)}{\Rightarrow}$ $\vec{\in}(K, \bar{\circ}(K, {}_{f_{19}, t_{19}}P)(\chi_8, \mu_2))$	μ_3	19	19	ensures the performing of certain commands
$K \overset{11(1)}{\Rightarrow} \vec{\in}(K, \mu_4)$ $\overset{11(2)}{\Rightarrow} \vec{\in}(K, \Diamond_4(\chi_6, \mu_3)) \overset{11(3)}{\Rightarrow}$ $\vec{\in}(K, \bar{\circ}(K, {}_{f_{20}, t_{20}}P)(\chi_6, \mu_3))$	μ_4	20	20	operator that ensures the performing of certain commands
$K \overset{4}{\Rightarrow} \vec{\in}(K, \wedge(\chi_{18}, \chi_{17}))$		26	26	logical expression
$K \overset{4}{\Rightarrow} \vec{\in}(K, \wedge(\wedge(\chi_{18}, \chi_{17}), \chi_{16}))$		27	27	certain logical expression
$K \overset{11(1)}{\Rightarrow} \vec{\in}(K, \mu_5)$ $\overset{11(2)}{\Rightarrow} \vec{\in}(K, \Diamond_1(\wedge(\wedge(\chi_{18}, \chi_{17}),$ $\chi_{16}), \chi_{20})) \overset{11(3)}{\Rightarrow}$ $\vec{\in}(K, \bar{\circ}(K, {}_{f_{30}, t_{30}}P)(-//-))$	μ_5	30	30	certain logical expression is true
$K \overset{11(1)}{\Rightarrow} \vec{\in}(K, \mu_5)$ $\overset{11(2)}{\Rightarrow} \vec{\in}(K, \Diamond_6(\chi_8, \mu_1')) \overset{11(3)}{\Rightarrow}$ $\vec{\in}(K, \bar{\circ}(K, {}_{f_{31}, t_{31}}P)(\chi_8, \mu_5))$	μ_6	31	31	ensures only if a certain logical expression is true

Each received image is added to the carrier of the performer . If such image already exists, then it can be redefined or extended by inheritance operation.

All images are "assembled" into a construction describing the branch operator, a definition derived from the written language construct-sentence.

In the same manner, a similar image described in the sentence 2 ("Conditional construct is an operator that allows performing certain actions if a certain condition is true") can be generated.

Let us construct the images corresponding to the considered sentence using the formula (4). We form the images of words $\kappa_{22} - \kappa_{37}$. The time and focus corresponding to this operation are given below (Table 3).

Table 3: Images of the words of sentence 2.

Image/word	Focus	Time	Image/word	Focus	Time
22/ conditional	32	32	30/certain	41	41
23/ consrtuct	33	33	31/action	42	42
24/ is	34	34	32/if	47	47
25/ an	35	35	33/a	48	48
26/ operator	36	36	34/certain	49	49
27/ that	38	38	35/condition	50	50
28/ allows	39	39	36/is	52	52
29/ performing	40	40	37/true	53	53

For all words of speech, we explicate the images-meanings χ_i, $i = \overline{22, 37}$and connect them with words using the formula (5). Let us consider the construction of composite images for the given sentence (Table 4).

For graphical representation of the performed operations and the structure of resulting constructions, let us construct the graphs (Fig. 1), vertices of which are the images (χ_i) corresponding to the words and the arcs are the images of relations ($\Diamond_i(\chi_m, \gamma_n)$, $\wedge(\chi_k, \chi_l)$).

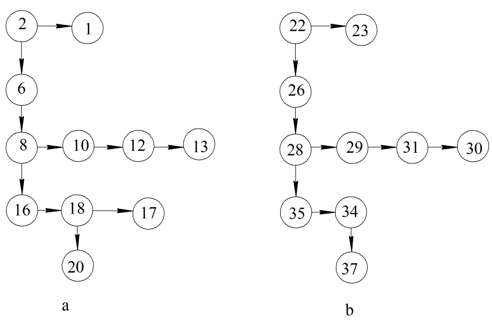

Figure 1: Structure of construction in the graph representation

Table 4: Constructing the composite image constructs of the sentence 2.

Conclusion	Image	Focus	Time	Prototype
$K \overset{4}{\Rightarrow} \vec{\in}(K, \wedge(\chi_{23}, \chi_{22}))$		33	33	conditional construct
$K \overset{11(1)}{\Rightarrow} \vec{\in}(K, \mu_7)$ $\overset{11(2)}{\Rightarrow} \vec{\in}(K, \diamondsuit_1(\wedge(\chi_{23}, \chi_{22}), \chi_{26}))$ $\overset{11(3)}{\Rightarrow} \vec{\in}(K, \bar{\circ}(K, {}_{f_{37}, t_{37}}P)$ $(\wedge(\chi_{23}, \chi_{22}), \chi_{26}))$	μ_7	37	37	conditional construct is an operator
$K \overset{4}{\Rightarrow} \vec{\in}(K, \wedge(\chi_{31}, \chi_{30}))$		43	43	certain actions
$K \overset{11(1)}{\Rightarrow} \vec{\in}(K, \mu_8)$ $\overset{11(2)}{\Rightarrow} \vec{\in}(K, \diamondsuit_8(\chi_{29}, \wedge(\chi_{31}, \chi_{30})))$ $\overset{11(3)}{\Rightarrow} \vec{\in}(K, \bar{\circ}(K, {}_{f_{44}, t_{44}}P)$ $(\chi_{29}, \wedge(\chi_{31}, \chi_{30})))$	μ_8	44	44	performing certain actions
$K \overset{11(1)}{\Rightarrow} \vec{\in}(K, \mu_9)$ $\overset{11(2)}{\Rightarrow} \vec{\in}(K, \diamondsuit_9(\chi_{28}, \mu_8))$ $\overset{11(3)}{\Rightarrow} \vec{\in}(K, \bar{\circ}(K, {}_{f_{45}, t_{45}}P)(\chi_{28}, \mu_8))$	μ_9	45	45	allows performing certain actions
$K \overset{11(1)}{\Rightarrow} \vec{\in}(K, \mu_{10})$ $\overset{11(2)}{\Rightarrow} \vec{\in}(K, \diamondsuit_{10}(\chi_{26}, \mu_9))$ $\overset{11(3)}{\Rightarrow} \vec{\in}(K, \bar{\circ}(K, {}_{f_{46}, t_{46}}P)(\chi_{26}, \mu_9))$	μ_{10}	46	46	operator that allows performing certain actions
$K \overset{4}{\Rightarrow} \vec{\in}(K, \wedge(\chi_{35}, \chi_{34}))$		51	51	certain condition
$K \overset{11(1)}{\Rightarrow} \vec{\in}(K, \mu_{11})$ $\overset{11(2)}{\Rightarrow} \vec{\in}(K, \diamondsuit_1(\wedge(\chi_{35}, \chi_{34}), \chi_{37}))$ $\overset{11(3)}{\Rightarrow} \vec{\in}(K, \bar{\circ}(K, {}_{f_{52}, t_{52}}P)(-//-))$	μ_{11}	52	52	certain condition is true
$K \overset{11(1)}{\Rightarrow} \vec{\in}(K, \mu_{12})$ $\overset{11(2)}{\Rightarrow} \vec{\in}(K, \diamondsuit_{12}(\chi_{28}, \mu_{11}))$ $\overset{11(3)}{\Rightarrow} \vec{\in}(K, \bar{\circ}(K, {}_{f_{53}, t_{53}}P)(\chi_{28}, \mu_{11}))$	μ_{12}	31	31	allows if a certain condition is true

8 Language realization of CSSt of human and community images

The result of realization of the structure (3) is the set of images of LC as a whole and its parts ${}_t\Omega(C_h({}_{\bar{k}_i}z_i))$:

$$ {}_t\Omega(C_h({}_{\bar{k}_i}z_i)) \supset ({}_t\Omega_s(C_h({}_{\bar{k}_i}z_i)) \bigcup {}_t\Omega_{pw}(C_h({}_{\bar{k}_i}z_i)) \bigcup {}_t\Omega_{aw}(C_h({}_{\bar{k}_i}z_i))), $$

where the sets ${}_t\Omega(C_h({}_{\bar{k}_i}z_i))$ – all the images formed by the performer ${}_{\bar{k}_i}z_i$ at the moment of time t, ${}_t\Omega_{pw}(C_h(z_i))$ – all images of the written LC, ${}_t\Omega_s(C_h(z_i))$ mean verbal constructions (including the ones corresponding to the ${}_t\Omega_{pw}(C_h(z_i))$), ${}_t\Omega_{aw}(C_h(z_i)))$ represent other images of the LC.

Constructions, ${}_t\Omega(C_h({}_{\bar{k}_i}z_i))$, the elements of communication inherent in a specific performer ${}_{\bar{k}}z_i \in Z$ will be called an individual language. Free language is a set of potentially possible constructions that any performer (individual) can recognize (understand) and use to transfer the information.

We assume that there is some subset $\bar{Z} \subseteq Z$ form n of the CSS performers C_h. The language of the community of performers $\bar{Z}$ will be considered as a set of constructions built on the structure (3) medium, as a result of its implementation: $L(t) = \bigcup_b({}_t\Omega^*(C_h({}_{\bar{k}_i}z_i)) \bigcap_t \Omega^*(C_h({}_{\bar{k}_j}z_j)))$, where $b = ({}_{\bar{k}_i}z_i,\ {}_{\bar{k}_j}z_j \in \bar{Z},\ {}_{\bar{k}_i}z_i \neq {}_{\bar{k}_j}z_j)$, ${}_t\Omega^*(C_h({}_{\bar{k}_i}z_i)) = {}_t\Omega_{pw}(C_h({}_{\bar{k}_i}z_i)) \bigcup_t \Omega_s(C_h({}_{\bar{k}_i}z_i)) \bigcup \Omega_{aw}(C_h({}_{\bar{k}_i}z_i))$. Language exists at some point in time t. LC belongs to the language if there are two or more of its carriers, capable of receiving and transmitting it $({}_t\Omega^*(C_h({}_{\bar{k}}z_i)) \bigcap {}_t\Omega^*(C_h({}_{\bar{k}}z_j)) \neq \emptyset,\ {}_{\bar{k}_i}z_i \neq {}_{\bar{k}_j}z_j)$. $L_{pw}(t) = \bigcup_b({}_t\Omega_{pw}(C_h({}_{\bar{k}_i}z_i)) \bigcap\ {}_t\Omega_{pw}(C_h({}_{\bar{k}_j}z_j)))$ is the written language of the community of performers, $L_s(t) = \bigcup_b({}_t\Omega_s(C_h({}_{\bar{k}_i}z_i)) \bigcap\ {}_t\Omega_s(C_h({}_{\bar{k}_j}z_j)))$ is the oral one.

The community includes groups of people which satisfy a certain relation of similarity. Presence of these groups allows distinguishing various sublanguages: languages of peoples, professional language, dialect languages, jargon, etc. Territorial characteristic, sphere of activity, standard of living, habitat, etc. can be used as an attribute on which a similarity relation is specified to distinguish the groups. For example, the professional language of programmers can be determined as $L_p(t) = \bigcup_b({}_t\Omega_p(C_h({}_{\bar{k}_i}z_i)) \bigcap\ {}_t\Omega_p(C_h({}_{\bar{k}_j}z_j)))$ where $b = ({}_{\bar{k}_i}z_i,\ {}_{\bar{k}_j}z_j \in \bar{Z},\ {}_{\bar{k}}z_i \neq {}_{\bar{k}}z_j,$ $occupation \hookleftarrow\!\rfloor\ z_i = occupation \hookleftarrow\!\rfloor\ z_j)$. For complete understanding and interaction of performers, it is desirable to draw a sample according to several characteristic, for example, occupation and territory (*location="Ukraine, Dnipro", occupation="programmer C#, senior"*), since in addition to generally accepted documented terms one can use definitions which represent, for example, transliteration or inexact translation of generally accepted English words. To select the professional LC of the chosen performers it is necessary to determine a set of images ${}_t\Omega_p(C_h({}_{\bar{k}_i}\bar{z}_i))$ as a set formed of elements with the same attribute for generalization ${}_t\Omega_p(C_h({}_{\bar{k}_i}\bar{z}_i)) = \{\omega_i : \forall\omega_i, \omega_j \in\ {}_t\Omega_{pw}(C_h({}_{\bar{k}_i}z_i)) \bigcup\ {}_t\Omega_s(C_h({}_{\bar{k}_i}z_i))\ \exists\omega_{pw,i} :$ $w_i \hookleftarrow\!\rfloor\ \omega_i = w_j \hookleftarrow\!\rfloor\ \omega_j\}$. For example, images cycle, variable, recursion have the common attribute-image programming by which they can be generalized to the programming term.

Carriers of the language and speech can be dynamic and static ones. Dynamic carriers can both store and generate constructions, for example, an individual, AI systems. Static carriers include those that cannot independently generate the language constructions; such carriers can be permanent books, audio discs, videotapes, and editable text computer files, soundtracks.

9 Analysis and identification of the similarity of language constructs

Working with static language carrier is useful in the tasks of information search and detection of plagiarism. To establish the fact of plagiarism, it is necessary to determine the matching content. To identify the matching content in the texts TXT_i and TXT_j it is necessary to distinguish the sets of printed word images of specific performers associated with this text $TXT_i \longrightarrow \Omega_{pw}(C_h({}_{\bar{k}_i}z_i))$ (the semantic content of some author's text), where $\longrightarrow$ is the display operation (can be implemented using s_{23}, s_{26}). Result of the operation $(TXT_i \longrightarrow \Omega_{pw}(C_h({}_{\bar{k}_i}z_i)))$ is a set of images $\{\omega_{pw,\,i}(C_h({}_{\bar{k}_i}z_i)) \in \Omega_{pw}(C_h({}_{\bar{k}_i}z_i)) : \forall txt_i \in TXT_i\, \exists txt_i \hookleftarrow\!\rfloor\, \omega_{pw,\,i}\}$, the attributes of which are the elements of the text $txt_i \in TXT_i$. The matching content (common fragments of texts) are defined as $(TXT_i \longrightarrow \Omega_{pw}(C_h({}_{\bar{k}_i}z_i)))\bigcap_b(TXT_j \longrightarrow \Omega_{pw}(C_h({}_{\bar{k}_j}z_j)))$, $b = ({}_{\bar{k}_i}z_i,\, {}_{\bar{k}_j}z_j \in \bar{Z},\, {}_{\bar{k}_i}z_i \neq {}_{\bar{k}_j}z_j)$. The display operation is performed by the same performers for two (and more) texts.

Let us construct the correspondence table on the basis of the relations' images and their semantic similarity (Table 5). The obtained correspondences are based on the semantic similarity of the concepts considered in the field of programming. As it can be seen from the Fig. 1 and Table 5, the constructs have structural similarity and some complete coincidences of parts.

The constructs mentioned in the table have incomplete correspondence ( ). To reveal the similarity of concepts, analysis of similarity is carried out by one and the same performer. Part of the images considered is identical, and they are expressed by the same language constructs. The other ones have certain semantic similarity (finding of the same depends on the level of the basic programming concepts knowledge of the model performer). Lets give some explanations for selected images and their constructs (in the lines of Table 5):

1. the branch operator is the means of implementing a conditional construct, i.e. one may talk of similarity of concepts (the latter is broader);

2. some commands imply certain actions which can be realized by the programming language;

3. performance of commands involves performance of actions; both precedents can be reduced to one concept and result;

4. provision (guarantee, assurance) represents more strict form of the permission;

5. the logical impression implies description of a certain condition;

6. similar ones, on the basis of two previous lines.

This approach can be used for the automated search for matching content in the tasks of anti-plagiarism of the natural language constructions-texts at the semantic level. The usage of image approach to the semantics representation allows reducing a few words to a single image-sense that solves the problem of synonyms in the anti-plagiarism systems.

Table 5: Correspondence of the image constructions.

#	Text 1		Text 2		Comment on the similarity of concepts
	Image/ construction	Text fragment	Image/ construction	Text fragment	
1	$\wedge(\chi_2, \chi_{21})$	branch operator	$\wedge(\chi_{23}, \chi_{22})$	conditional construct	In fact, the same name found in different literature sources
2	$\wedge(\chi_{13}, \chi_{12})$	certain commands	$\wedge(\chi_{31}, \chi_{30})$	certain action	Commands and actions are close concepts, because they are realized by the programming language operators
3	μ_2	performing of certain commands	μ_8	performing certain actions	Participial construction with similar meaning
4	μ_4	operator that ensures the performing of certain commands	μ_{10}	operator that allows performing certain actions	Participial construction with similar meaning
5	μ_5	certain logical expression is true	μ_{11}	certain condition is true	Logical expression describes a condition
6	μ_6	ensures if a certain logical expression is true	μ_{12}	allows if a certain condition is true	It imposes the same condition on execution of an action

10 Conclusions

The constructed model of the NL is based on the figurative perception of the world by an individual. The model basis is represented by the formal grammars, which is widely recognized method of calculations. Formalization is provided for the thinking processes which are inextricably connected with encoding and transmission of thoughts using the elements of communication, such as gestures, facial expressions, speech, and writing. The latter are the basis for determining free and individual languages of people which are relevant to the concepts of the objective and subjective languages [20].

The presented model, in contrast to well-known ones:

- uses a single constructive approach for modeling all components and operations;
- is applicable to different forms of presentation of the language constructs;
- covers various aspects (syntactic, semantic ones) of the language;
- is closer to the natural processes;
- unlike the models that collect statistics, construct matrices, etc., the observed model already has a basis, i.e. the extensible, dynamic carrier of the performer, on the elements of which the relations are constructed and operations are performed;
- in contrast to the models used in NLP, for example, n-gram ones, it uses the meaning, not numeric attributes.

The model makes it possible to:

- consider NL as a set of communicative abilities of an individual which takes into account his/her language features and a person performer of the given model;
- consider the language as a constructive process which can be used as the basis for creating a methodology for building the systems with high degree of intellectuality;
- formally represent classifications of the language (areal classification, classification by the sphere of use (common, professional)), taking into account the characteristics of its speaker/carrier;
- improve the semantic NLP, in particular, in the tasks of comparing and identifying matching semantic content in texts, thus significantly reducing the influence of synonyms, homonyms, paraphrases, and translation.

The scope of the presented model covers NLP-components of robots and applications, including the systems of translation and anti-plagiarism, as well as expert systems.

References

[1] Beaufort, Richard, Roekhaut, Sophie, Cougnon, Louise-Amélie, and Fairon, Cédrick. A hybrid rule/model-based finite-state framework for normalizing sms messages. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 770–779. Association for Computational Linguistics, 2010.

[2] Bengio, Yoshua, Ducharme, Réjean, Vincent, Pascal, and Jauvin, Christian. A neural probabilistic language model. *Journal of Machine Learning Research*, 3:1137 – 1155, 2003.

[3] Bisikalo, O. V. Conceptual model of imagery analysis and synthesis of natural language constructions. *Mathematical Machines and Systems*, (2):184 – 187, 2013.

[4] Bisikalo, O. V. *Formal methods imagery analysis and synthesis of natural language constructions: monograph.* Vinnitsa: VNTU, 2013.

[5] Brown, Peter F, Desouza, Peter V, Mercer, Robert L, Pietra, Vincent J Della, and Lai, Jenifer C. Class-based n-gram models of natural language. *Computational Linguistics*, 18(4):467 – 479, 1992.

[6] Cambria, Erik and White, Bebo. Jumping nlp curves: a review of natural language processing research. *IEEE Computational Intelligence Magazine*, 9(2):48 – 57, 2014.

[7] Choudhury, Monojit, Saraf, Rahul, Jain, Vijit, Mukherjee, Animesh, Sarkar, Sudeshna, and Basu, Anupam. Investigation and modeling of the structure of texting language. *International Journal on Document Analysis and Recognition*, 10(3-4):157 – 174, 2007.

[8] Chowdhury, G. Natural language processing. *Annual Review of Information Science and Technology*, 37(1):51–89, 2003.

[9] Eger, Steffen, Schenk, Niko, and Mehler, Alexander. Towards semantic language classification: Inducing and clustering semantic association networks from europarl. In *Proceedings of the Fourth Joint Conference on Lexical and Computational Semantics*, pages 127 – 136, 2015.

[10] Franco-Salvador, Marc, Gupta, Parth, and Rosso, Paolo. Cross-language plagiarism detection using a multilingual semantic network. In *European Conference on Information Retrieval*, pages 710–713. Springer, 2013.

[11] Kohan, Ya. O. On the possibilities of formalizing natural languages. In *TAAPSD*, volume 3, pages 137 – 143, 2016.

[12] Krak, Iu.V., Barmak, O.V., and Romanyshyn, S.O. The method of generalized grammar structures for text to gestures computer-aided translation. *Cybernetics and Systems Analysis*, 50(1):116 – 123, 2014.

[13] Krak, Iu.V., Kuznetsov, V.A., and Ternov, A.S. Facial expressions modeling based on a parametric models of a human head. *Artificial Intelligence*, pages 154 – 170, 2013.

[14] Le, Quoc and Mikolov, Tomas. Distributed representations of sentences and documents. *ICML*, pages 1188 – 1196, 2014.

[15] Liddy, Elizabeth D, Paik, Woojin, and Yu, Edmund Szu-li. Natural language processing system for semantic vector representation which accounts for lexical ambiguity, February 1999. US Patent 5,873,056.

[16] Marchenko, OO. A method for automatic construction of ontological knowledge bases. i. development of a semantic-syntactic model of natural language. *Cybernetics and Systems Analysis*, 52(1):20–29, 2016.

[17] Mazzei, Alessandro and Lombardo, Vincenzo. Building a large grammar for italian. In *LREC*, 2004.

[18] Micol, Daniel, Muñoz, Rafael, and Ferrández, Óscar. Investigating advanced techniques for document content similarity applied to external plagiarism analysis. In *Proceedings of Recent Advances in Natural Language Processing*, pages 240 – 246, Hissar, Bulgaria, September 2011.

[19] Mohri, M. On some applications of finite-state automata theory to natural language processing. *Natural Language Engineering*, 2(1):61 – 80, 1996.

[20] Nekipelova, I. M. Zarifullina, E. G. Language system as a natural multi-level classification of a high degree of reliability. *Modern studies of social problems*, (6 (26)), 2013.

[21] Nivre, Joakim, de Marneffe, Marie-Catherine, Ginter, Filip, Goldberg, Yoav, Hajic, Jan, Manning, Christopher D, McDonald, Ryan T, Petrov, Slav, Pyysalo, Sampo, Silveira, Natalia, et al. Universal dependencies v1: A multilingual treebank collection. In *LREC*, 2016.

[22] Osman, Ahmed Hamza, Salim, NAOMIE, Binwahlan, S, Hentabli, H, and Ali, ALBARAA M. Conceptual similarity and graph-based method for plagiarism detection. *Journal of Theoretical and Applied Information Technology*, 32(2):135 – 145, 2011.

[23] Schrader, Yu. A. *Equality, resemblance, order*. M .: Publishing house Science, 1971.

[24] Shynkarenko, V. and Vasetska, T. Modeling of ranking alternatives methods of analysis hierarchy of means of constructive-synthesizing structures. *Mathematical machines and systems*, (1):39 – 47, 2016.

[25] Shynkarenko, V. I. Qualification characteristics of natural, artificial and hybrid intelligence. *Artificial intelligence*, 67(1 – 2):9 – 19, 2015.

[26] Shynkarenko, V. I. and Ilman, V. M. Constructive-synthesizing structures and their grammatical interpretations. part i. generalized formal constructive-synthesizing structure. *Cybernetics and Systems Analysis*, 50(5):665 – 672, 2014.

[27] Shynkarenko, V. I. and Ilman, V. M. Constructive-synthesizing structures and their grammatical interpretations. part ii. refining transformations. *Cybernetics and Systems Analysis*, 50(6):829 – 841, 2014.

[28] Shynkarenko, V. I. and Kuropiatnyk, O. S. Constructive productive model of graph representation of text. *Problems of Programming*, (2 – 3):63 – 67, 2016.

[29] Shynkarenko, V. I. and Zabula, G. V. Constructive model of adaptation of data structures in operational memory. part i. constructing texts of programs. part ii. designers of scenarios and adaptation processes. *Science and Transport Progress. Bulletin of Dnipropetrovsk National University of Railway Transport*, 61, 62(1, 2):109 – 121, 88 – 97, 2016.

[30] Shynkarenko, V.I., Ilman, V.M., and Skalozub, V.V. Structural models of algorithms in problems of applied programming. i. formal algorithmic structures. *Cybernetics and Systems Analysis*, 45(3):329 – 339, 2009.

[31] Silberztein, Max. *Formalizing Natural Languages: The NooJ Approach.* John Wiley & Sons, 2016.

[32] Silberztein, Max. A new linguistic engine for nooj: Parsing context-sensitive grammars with finite-state machines. In *International Conference on Automatic Processing of Natural-Language Electronic Texts with NooJ*, pages 240–250. Springer, 2017.

[33] Tsatsaronis, George, Varlamis, Iraklis, Giannakoulopoulos, Andreas, and Kanellopoulos, Nikolaos. Identifying free text plagiarism based on semantic similarity. In *Proceedings of the 4th International Plagiarism Conference*, 2010.

Received ...