

TYPES OF LINGUISTIC MODELS AND THEIR APPLICATION TO SOLVING OF LINGUISTIC TASKS

© M. O. Lukach, 2013

This paper outlines understanding of the notions ‘model’ and ‘modeling’ in linguistics. The main linguistic models and their application to solving of different linguistic tasks are described.

Keywords – modeling, model, generative model, research model, analytical model, synthetic model.

Introduction. General statement of the problem

Over the last decades we observe the rapid development of applied linguistics due to the permanently increasing need for application of natural language mechanisms to information and computer human-machine systems [3, p. 3].

Modeling of natural language and speech became the issues of particular importance. It encourages researchers to develop a variety of linguistic models that could solve practical problems of linguistics, namely, information retrieval, machine translation, natural language understanding, knowledge extraction and localization from text et al.

Analysis of recent studies and publications

One of the important linguistic models is model ‘Meaning \Leftrightarrow Text’ developed by I. A. Melchuk.

Although this multilevel model of transformation from meaning into text and vice versa was established already in the second half of the twentieth century, but it still continues to be investigated and is applied to the problems of machine translation. In addition, the markup of syntactic sub-corpus of Russian National

Corpus is performed using linguistic processor ETAP-3, based on the principles of model ‘Meaning \Leftrightarrow

Text’ [8].

We should also mention the grammatical model HPSG (Head-driven phrase structure grammar) - generative grammatical theory, developed by C. Pollard and I. Sag in the early 1980s [9, p. 44]. Since that time, different parsers based on HPSG formalism have been written. The HPSG grammars of different languages are being developed within the project (DELPH-IN).

1. The notions ‘model’ and ‘modeling’ in linguistics

In the linguistic literature, the term ‘model’ was first used by American Z. Harris in 1944 when he described the difference between the methodological techniques of the two linguists S. Humen and E. Sapir, and then in 1951 he used this term again to indicate the results of E. Sapir’s descriptive terminology. But more specifically (i.e., for use in grammar) the term ‘model’ was used by C. Hockett in 1954 and by N. Chomsky in 1956. The latter ones determine a **model** as a generalized and formalized structure or process of one or another facts of language. In 1957 A. Oettinger already talks about models

not only as reproduction of reality, but also as an impact on this reality. Another linguist V. Yngve extends the notion of model from the specific languages to the mechanism of language on the whole [5, p. 14].

The term ‘model’ is widely used in various fields of knowledge and, therefore, its definitions differ depending on the particular application. Several approaches to understanding of this term in linguistic science will be outlined next.

Model (French modèle, from Latin modulus - a measure) in linguistics - 1. an artificially created by a linguist real or imaginary device that reproduces, imitates with its behavior (usually in a simplified form) the behavior of some other (“real”) device (the original) for linguistic purposes; 2. a sample, which serves as standard (benchmark) for mass reproduction, the same as ‘type’, ‘diagram’, ‘paradigm’, ‘structure’, etc. (e.g., ‘a model of conjugation or declination’, ‘a word formation model’, ‘a model of sentence’, etc.) [4, p. 304].

In another dictionary [2, p. 238] a similar definition occurs. **Model** (English pattern, model) - 1. (scheme) regular placement of sequences of segment and above-segment elements in complex linguistic units, characterized with properties of language structure and regularly reproduced in speech (e.g., a model of a sentence, a word formation model); 2. peculiar to the given language paradigmatic scheme of elements (components) of its structure, compact symbolic representation of this scheme (or any of its parts or elements), which is the basis for modeling a language as one of the methods of its study (e.g., generative model).

Modeling – a scientific method which consists in the schematic reproduction of an object that cannot be directly observed or is very difficult [2, p. 238]. The need for modeling arises in all those scientific fields where the object of science is inaccessible to direct observation. In such cases it is usually likened to some “black box”, and the only thing we know about it is that it receives the raw materials as an “input” and produces the end products as an “output”. The task, therefore, is to learn the content of the “black box” - that mechanism, hidden from the researcher that performs processing of the initial materials into the end products. Since it is impossible to disassemble the “black box” without breaking its functioning, there is only one way to know the object: on the basis of comparing the initial and final data, an image of the object should be constructed, i.e., the hypothesis of its possible structure should be formulated and implemented as a logical machine, that can process a certain material in the same way as it does the “black box” itself. If our logic device works like this, it is an approximation or a model of the object, and we can assume, that the mechanism, laid in it, in all essential details coincides with the mechanism, contained in the “black box”. As a result, we can understand the mechanism unavailable to our observation, comparing it to the mechanism, the structure of which is well known. Thus, as noted by Yu. D. Apresyan, the meaning of **modeling** is to examine, instead of the object’s properties hidden from us, those properties of the model that are specified explicitly, and to distribute to the object all those laws that were derived for a model [1, p. 78-79].

Another Russian linguist [6, p. 8] defines the notion of ‘**modeling a language**’ as a method by which, first, based on the most common features of specific languages, some hypotheses about the structure of a language as an abstract semiotic system are formulated, and then the relation between the results of these hypotheses and the facts of the real languages, described with the specific linguistic disciplines, is set.

There are the following basic properties of the models [1, p. 79-99]:

1. Only those phenomena can be modeled, essential properties of which are limited to their structural (functional) characteristics and in no way are related to their physical nature. As a model of an object for which only its functional properties are significant, any device that is functionally similar to it should be considered. In other words, the behavior of the model should be similar to the behavior of the object; the material in which it is implemented may be different (and practically it is always different) from the material in which the object is constructed.

All other properties of the model are related to its main characteristic, namely, that it is a functional approximation of an object.

2. A model is always a certain idealization of an object. In order to understand the complex phenomena, one should begin with the study of the simplest and common cases, even if they never occur in pure form, and from them move to more complex and specialized cases. The most important example of an idealization of a linguistic object is the assumption that the number of sentences in the language is infinite and that the length of sentences, in principle, has no limits (i.e., it is possible that there are sentences, the length of which is greater than any previously specified one). Obviously, the actual number of written or spoken sentences in any language is great, but finite, however, to explain the speaker's ability to build new, never spoken or written before sentences one should not consider this actually observed situation, but some ideal situation instead.

3. Typically, the model operates not with notions about the real objects but with notions about the ideal objects so-called constructs, not derived directly and unambiguously from the data, but built "freely" based on some general hypotheses, suggested by a set of observations and researcher's intuition. Any model is a structure logically derived from hypotheses, using some mathematical apparatus. Ideally, linguistic constructs are built without direct reference to phonetic or semantic substance of those phenomena, for studying of which they were created.

4. Any model must be formal. The model is considered to be formal if the initial objects, that connect their confirmation and rules of handling them (rules of formation or selection of new objects and confirmations), are set in it unambiguously and explicitly. Ideally, any formal model is a mathematical system, so in a sense, the notion of formality is equivalent to the concept of mathematical precision or single-valuedness. A formal model is associated with the data with the help of one or another interpretation. To make an interpretation of a model means to specify the rules, probabilistic or strict, the substitutions of objects of some subject areas, e.g., language, instead of objects (characters) of the model.

5. Any interpreted model should have property of explanatory power. The model has this property if it 1) explains the facts or data of specifically conducted experiments that cannot be explained in terms of the old theory 2) provides previously unknown, but fundamentally possible behavior of an object, which was later confirmed with new observations or experiments. In both cases, the explanatory power of the model increases with the fuller extent of coincidence of predictions with experimental data.

As a basic concept of any language model the notion of text in a certain alphabet can be considered. The text should be understood as actually uttered text ("a set of correct sequences of language") or as some its generalization, e.g., the sequence of classes that correspond to the elements of the real text [7, p. 17].

Building a model involves the following steps [1, p. 99]:

1. Recording of facts that require explanation.
2. Formulation of hypotheses to explain the facts.
3. Implementation of the hypotheses in the form of models that not only explain the initial facts, but also foresee new facts that were not previously observed.
4. Experimental verification of the model.

2. The main types of linguistic models

The types of linguistic models are considered on the basis of different classifications. Thus, according to the nature of the object under consideration, the linguistic models are subdivided into: [1, p. 99-100]

1. The models which consider as an object the specific linguistic processes and phenomena that imitate a human's speech activity. They were first developed by the Prague structuralists.

2. The models which consider as an object the procedures leading to the discovery of one or another linguistic phenomenon. These models, to some extent, imitate the research activity of a linguist and are called research models. The first serious step in their development was made by American descriptivists that tried to create the universal algorithm of discovering the grammars of natural languages.

3. The models which consider as an object ready linguistic descriptions, and not human's speech activity or research activity of a linguist. If regular grammar is a theory of specific language processes,

then the model of the third type is a theory of a theory or metatheory. Such types of models were first developed by the glossemantics school.

These metamodels include the system of criteria and theoretical evidences (metalanguage), with the help of which out of several alternative models that simulate the same phenomenon, the best one can be chosen. In computational linguistics, the metamodels (metatheorists), which objects are not separate linguistic concepts, but coherent language models, are developed [4, p. 304].

Depending on what is meant by the notion “to know a language” (whether or not the characteristic of knowing the meaning of the words is included in it), the models of speech activity can be divided into: 1) non semantic or purely syntactic, that imitate knowing of grammar, i.e., the ability of native speakers to understand and construct grammatically correct, but not necessarily meaningful phrases, 2) semantic, that imitate the ability of native speakers to understand and build meaningful sentences. Depending on which side of speech activity - listening or speaking - is the subject of modeling, the models of speech activity are divided into models of analysis and models of synthesis. The model of analysis is a finite number of rules that can analyze an infinite number of sentences of a given language. The syntactic analytical models receive text as an “input” and produce for each sentence the record of its syntactic structure as an “output”. The semantic analytical models receive the same material as an “input” and produce the meaningful record (sense image) of each sentence in a special semantic language as an “output”. The model of synthesis is a finite number of rules that can build an infinitely large number of correct sentences. The syntactic synthetic models use the record of syntactic structure of sentences as an “input” and give the correct sentences of a given language as an “output”. The semantic synthetic models receive the meaningful record of some sentence in a special semantic language as an “input” and produce a set of sentences of natural language, synonymous with this sentence as an “output” (Table 1) [1, p. 106-107].

Table 1

The main types of linguistic models

Characteristics	What linguist knows	Kind of input information	Kind of output information	Goal
Type of model				
Research models	Text (and set of correct phrases)	Text	Grammar and vocabulary	To model linguist's activity
Analytical models	Grammar and vocabulary	Text	Representation of text structure	To model understanding of text
Synthetic models	Grammar and vocabulary	Representation of text structure	Text	To model creation of text
Generative models	Grammar and vocabulary	Alphabet of symbols and rules of creation and transformation of phrases	Set of correct phrases and representation of their structure	To model the ability of distinguishing the correct and incorrect in language

The notions of analytical and synthetic models are related to the notion of reversibility of the model. Model M1 is considered as reversed to the model M2 if the initial objects of M1 are the finite objects of M2 and the finite objects of M1 - the initial objects of M2. Some researchers consider the synthetic models as reversed to the analytical ones. This approach eliminates the need to build two separate models of the given phenomenon: an analytical model can be obtained using a simple reference to the synthetic one or vice versa. Except for the models of analysis and synthesis, there are also so-called generative models which are, in a sense, intermediate between the models of analysis and synthesis. Generative model is a construction that contains an alphabet of characters and a finite number of rules of formation (and transformation) of expressions out of the elements of this alphabet and can construct an infinite number of correct sentences of the given language and assign to each of them a certain structural characteristic [1, p. 106-107].

Depending on which mathematical form a model is represented, the models are divided into calculuses and algorithms. Calculus is a system of permits (permit), and an algorithm is a sequence of orders (commands). Usually calculus is a mathematical system that includes: 1) initial (primary or undefined) notions, which names form the “alphabet of symbols”; 2) primary (not proven) statements about connections between these concepts (axioms); 3) rules of deriving new statements (theorems) out of already existing ones. Sometimes instead of axioms and rules of inference, the rules of creation and transformation of expressions out of the alphabet’s elements are used. In calculuses the so-called recursive definitions and rules are commonly used. The definitions and rules are considered as recursive if they are built in two steps, the first step comprises determination of the simplest separate case, and the second one – determination of the overall case out of a separate one. Calculus allows you to specify, with the help of the finite apparatus, all objects of a given set, including an infinite one (e.g., all sentences of the given language). This property of calculuses should be used in linguistics which deals with very large or infinite number of units. An algorithm is a sequence of commands, the implementation of which leads to the selection (or building) of the desired object. The algorithm should allow completely automatic implementation. The algorithm written in a machine understandable language is a program. Any model, including calculus, should be presented in the form of an algorithm (or provided with an algorithm) in order to be implemented on machine because machine understands only language of commands and not language of permits. Research, analytical and synthetic models are usually made in the form of algorithms and generative models are commonly used in the form calculuses [1, p. 108-110].

Depending on what kind of rules are used in the model, there are probabilistic (statistical) and deterministic (structural) models. There are also mixed structural and statistical models. Natural languages in most cases are arranged in such way that a small number of rules covers the basic set of facts, but in order to explain a few remaining facts, mostly unproductive, a very big number of rules is required. So sometimes it is more advantageous to explain a certain set of facts not with deterministic model, which due to the large number of rules can be extremely unwieldy for performing quite definite task, but with probabilistic model, which requires fewer number of purely statistical rules and, therefore, is less unwieldy. The loss of accuracy is compensated in such model with its relative simplicity. The most important deterministic models are models of binary differential structures in phonology and morphology, the model of immediate constituents, transformational and applicational models in the syntax, the model of “semantic factors” in the field of semantics [1, p. 112].

Formulation of objectives

The purpose of the article - to outline the approaches to definition of the notion ‘model’ in linguistics, as well as to review the main types of linguistic models and their application to solving of different linguistic problems.

Statement of the main material

1. The research models

The research models can be divided into three classes, depending on what information is used in them as initial. In the model of the first class only text is used as initial information, and all information about the system, i.e., language that generates this text, is extracted from text data only. This is an example of classic decoding models. In the models of the second class not only text, but also a set of correct sentences of the language are given. This means that during the creation of a model a linguist gets help from an informant that has to say whether each proposed sentence is correct or not. The linguist himself can be an informant if he is fluent in the given language. Finally, in the models of the third class not only a text and a set of correct phrases, but also a set of semantic invariants are given. This means that the informant must not only determine the correctness of each proposed sentence, but also he should be able to say about any two phrases whether they mean the same or not [1, p. 101].

1.1. The decoding research models

The first type of the research models is called models of linguistic decoding as they, in addition to their primary theoretical function (study of linguistic notions and statements), may have, in principle, some applicative - decoding function. The initial information for this type of models is a text which is unknown in advance. We know neither the language (code) used for “encryption” of the text, nor the genetic connections of this language with already known languages, nor translation of the text into already known languages, nor an area of reality, which is described by the text. All other information, i.e., information about the elementary units of the text (letters or sounds, morphemes, words, sentences, and, finally, senses), classes of the elementary units (vowels and consonants, lexical and grammatical morphemes, parts of speech, types of sentences, semantic fields) and laws of combination of items of different classes (e.g., syntactic relations of words in a sentence) should be completely automatically derived from the text [1, p. 120].

The examples of such models are algorithms of B. V. Sukhotin and Z. Harris. Some general features of these algorithms will be outlined next. First, in the basis of all such algorithms lie some simple and common understandings of a language, confirmed by certain universal laws, for example, “a letter is a stable combination of dots”; “a morpheme is a stable combination of phonemes”; “a word form is a stable combination of morphemes”; “in every natural language there are at least two levels - the level of meaningful units (morphemes, word forms, constructions) and the level of meaningless units (phonemes)”; “in any language there are lexical morphemes, and the distribution of lexical morphemes in a text differs from the distribution of grammatical morphemes”; “the syntactic differences correspond to the semantic differences”; “the words with similar meaning are close to each other in a text”, etc. Secondly, all such algorithms use information on the distribution of the elements and their numerical parameters. Thirdly, usually, first, a set of feasible solutions is defined, and then in this set the best solution is found using the so-called utility functions. The utility functions are numeric functions, which get a certain (e.g., minimum possible or maximum possible) value if the correct (best) decision was made. Each utility function formalizes some meaningful hypothesis about the possible properties of the object under study. The algorithm of B. V. Sukhotin for the allocation of vowels and consonants assumes that the alphabet of letters, with the help of which the given text was recorded, is already defined. Its task is to provide the key to the rules of reading the text, without which the decoding cannot be considered as complete. Any division of the alphabet into two classes is considered as an acceptable solution. The algorithm formalizes the following substantial hypothesis about the properties of vowels and consonants: the vowels and consonants are interchanged in a text; there are no texts which consist only of vowels, as there are no texts which consist only of consonants; vowels are followed by consonants and consonants are followed by vowels [1, p. 122-123].

Another example of the decoding research model is the morphological algorithm of Z. Harris. The algorithm works in such way that, first, from the text, recorded in the form of phonological transcription, some sentence is selected, e.g., an English sentence [Hiyzklever] – “he is clever”. Then, all sentences that begin with the same phoneme [h] are found, and the number of different phonemes that follow the [h] is calculated in these sentences. These phonemes are called “successors” of [h]. The number of “successors” of [h] is 9. After that, all sentences that begin with the same two phonemes, as in the given sentence, i.e., [hi] are found, and the number of “successors” of this pair of phonemes is counted (which is 14), etc. In different places of the sentence the number of “successors” can change: it grows forming peaks and then falls. Assuming that the assumption, that phonological diversity at the junction of morphs is higher than in the middle of the morphs, is correct, the peaks should occur at the boundaries between the morphemes (Table 2).

Table 2.

The example of the application of the morphological algorithm of Z. Harris

h	i	y	z	k	l	e	v	Θ	r
9	14	29	29	11	7

In this sentence there are two peaks - after the phoneme [y] and after the phoneme [z], and exactly in these areas there should be morphological boundary. The resulting segmentation (he/is/clever) reflects the actual situation [1, p. 127-128].

1.2. The experimental research models

The experimental research models pursue the same goal as the decoding ones: the task of transition from the text to the “system” is also assigned to them, that is the task of receiving from the text data the information about the elementary units of language, the classes of the elementary units, the laws of combination of the elements of the different classes, etc. The difference between these types of models lies in the method of obtaining the generated information and in the nature of the initial information. If in the decoding models a linguist limits himself to pure observation, capturing the behavior of the given object (phonemes, morphemes, constructions) in natural conditions of the text and exposing the observations only to mathematical processing, then in the experimental research models he observes the material in artificial conditions, exposing it to all kinds of transformations. In other words, in the models of the second type he uses an experiment which is not possible within the decoding models. In order to be able to conduct an experiment it is not enough to have only text at linguist’s disposal, because every real text has limited extent and, therefore, does not contain all principally possible phrases in the given language. In addition to the text, the linguist must have a set of correct sentences, in principle infinite, of the given language. In other words, in order to create an experimental model the linguist should be able to address to the informant, who should say about every sentence, whether it belongs to a number of phrases in this language or not (is it possible to say like that or not) [1, p. 149].

The example of this type of models is the model of immediate constituents, developed by L. Bloomfield, R. Wells, Z. Harris as an experimental research model, but which subsequently was widely used in the modeling of speech activity.

As we can see, unlike decoding research models, the experimental research models are more flexible and with them you can get more information about units of language from the original text.

2. The models of speech activity

In general, the models of speech activity are divided into generative models, syntactic models of analysis and semantic models of analysis.

2.1. The generative models

The generative model, which is usually presented in the form of calculus, is a finite set of rules that can define or generate all correct, and only correct, objects of a certain set, including the infinite one, and assign to each object a specific analysis [1, p. 198]. There are three types of syntactic generative models:

1. The immediate-constituent generative model

The example of the immediate-constituent generative model is a model of V. Yngve. The model consists of 1) grammar and 2) mechanism. Grammar – is the finite irregular set of rules on the immediate constituent expansion of the following types: a) $A \rightarrow B + C$ (the symbol in the left part of the rule is expanded into two new characters, e.g., $S \rightarrow NP + VP$, where NP - subject group and VP – predicate group. b) $A \rightarrow B$ (the symbol in the left part of the rule is replaced by another one, usually a terminal symbol, i.e., a symbol to which any other rules of immediate constituent grammar are not applied), e.g., $T \rightarrow the$, a (definite or indefinite article). c) The third group consists of rules of the type $B \rightarrow D + \dots + E$, designed to generate structures with so-called discontinuous constituents, i.e., constituents such as “if ... then”, between parts of which there may be other components. The second part of the model is mechanism - an idealized but physically implemented computing machine, that consists of three interrelated parts: output device (prints output elements one by one), decision device (contains a single character that is currently expanding

according to one of the above described rules), permanent memory (stores all grammar rules, i.e., the rules on the immediate constituent expansion), high-speed memory (stores intermediate results) [1, p. 199-201].

2. Transformational generative model

An adequate grammatical model must contain several levels of description appropriate to the levels of language structure, and each level of description must represent a specific set of rules. Transformational generative grammar comprises three levels of description, each of which is defined as partially ordered set of rules of a certain type and is one of the components of the transformational generative grammar. Among such components there are: 1) the rules of the immediate constituents, 2) transformation rules, and 3) morphonological rules. The initial rules for transformational generative grammar are rules on the immediate constituent expansion. They are subdivided into context-free (always used) and context-dependent (used only if the left character of the rule is in a particular context). These rules are the substitution rules of the type $A \rightarrow B + C$, on which the following restrictions are imposed [1, p. 213-216]:

1) Each rule is applied to a particular string no matter how this string was received. Thus, the same rules of expansion of symbol NP (noun group) apply in case if NP was obtained from the rule $S \rightarrow NP + VP$ and if NP was obtained from the rule $VP \rightarrow V + NP$. The only prerequisite for the application of the given rule is presence of the immediate constituents of a symbol from the left part of this rule in a string.

2) As a result of applying the rules of immediate constituents to the original string there must be a tree of components, as transformation is applied to trees rather than strings, and the tree of components must be attributed to this sentence explicitly. Therefore, in the immediate constituent component of the transformational grammar it is forbidden to expand more than one character simultaneously.

3) Any recursive rules, i.e., rules of the type $A \rightarrow A$ or $A \rightarrow A + B$, are prohibited. Recursive rules should be moved to the transformational component of grammar.

4) The rules of the type $A + B \rightarrow B + A$ (permutation) can be interpreted as two recursions and, therefore, should be removed from the immediate constituent component of the transformational grammar. These rules should also be transferred to the transformational component of the grammar.

To the advantages of the transformational model the following facts should be attributed: 1) transformational calculus provides generation of such types of sentences that cannot, probably, be derived from the rules of the immediate constituents; 2) transformational deriving of some complex syntactic types from the simple ones has the advantage of describing the generative process and its results more naturally; 3) transformational calculus allows to establish quite deep connections between the syntactic structure of the sentence and its meaning which are not fixed according to the rules of the immediate constituents [1, p. 223].

3. Applicational model.

An applicational model is fundamentally different from the above mentioned generative models in two aspects: its logical structure and linguistic concept, formalized in it. The peculiarity of the applicational model's logical structure lies in the fact that the generative process is explicitly represented in it as the one that runs on two significantly different levels - level of constructs and level of observation. The generative process begins with the output of ideal objects - constructional analogs of words and sentences that in the second stage of the generative process are transformed by certain rules of interpretation into real words and sentences of one or another natural language. The generative mechanism, which products are ideal objects, operates independently of the rules of interpretation. The ideal objects do not contain information about grammatical (word creation and inflection) categories of this or that particular language, such as, gender, number, case, person, time, aspect, etc. These and similar categories occur only during the interpretation of the model. Therefore, either the mechanism itself with its ideal words and rules of creation of ideal sentences, or generated by it set of ideal objects can be considered as an ideal language that can be used, in particular, as an intermediary language during typological comparison [1, p. 223].

Unlike the immediate-constituent model and transformational model, which objects of operations are strings or trees, that are converted into new strings or trees, in the applicational generative model there are

two types of linguistic objects - classes of words and complexes of words. In accordance with this, there are two different, though related to each other, generative mechanisms in transformation model – the device of generation of words classes and the device of generation of words complexes. The first one models the relations between units on the paradigmatic axis of language, and the second one - the relations between the units on the syntagmatic axis of language. In linguistic view, the applicational generative model differs from transformational model and immediate-constituent generative model in one more aspect. It uses two different operations: application, which is the only rule for objects creation and transformation, which is the only rule for their invariant conversion [1, p. 224].

An alphabet of characters, used in the applicational generative model, consists of four classes of characters of ideal stems: N (interpreted as class of stems of non-derivative nouns, such as, house, fish, window), V (interpreted as class of stems of non-derivative verbs, such as, go, ride), A (interpreted as class of stems of non-derivative adjectives, such as, white, large) and D (interpreted as class of stems of non-derivative adverbs, such as, there, then, that way). Out of these characters, by the rules of application, the complexes are derived. The characters N, V, A, D, and any n-tuple (pair, three, four or five) of these characters, formed by a single use of any combination of application rules to symbols N, V, A, D., such as: big house (AN), overcast day is seen (ANV) are considered as an elementary complex [1, p. 225].

The examples of the applicational generative model are: model of N. Chomsky and model S. K. Shaumyan based on the two-stage principle.

Thus, compared with the immediate-constituent generative model, transformational generative model has three levels of description that can deeper reflect structure of language. Applicational generative model, in contrast to the two previous models, has different linguistic background, meaning it, with the help of rules of application, generates classes of words and complexes of words out of the alphabet of characters of four classes of ideal stems.

2.2. The syntactic models of analysis

There are four main approaches to solving the problem of automatic text analysis (transition from text to syntactic information contained in it): sequential analysis, predictive analysis, fulcrum method and filter method.

The basic principles of sequential automatic method of text analysis were formulated by V. Yngve in 1954. A device that analyzes the text should store in the memory a list (dictionary) of typical syntagmas of the given language, written in the form of sequences of word classes. Syntagmas are considered as benchmarks; the syntactic structure of a sentence is searched by machine as a result of sequential comparison of different strings of word forms of the given sentence with the benchmarks that are stored in the dictionary, and each time the longest syntagma is selected as a term for comparison. The found phrases get numbers of the corresponding syntagmas. After identifying all phrases of the given sentence the relations between them are established (the hierarchy of phrases), and syntactical analysis of the sentence is considered to be concluded. The most consistently described approach to solving the problem of automatic text analysis is given in the algorithm of I. A. Melchuk. The procedure for the text analysis is divided into two large parts – the preliminary preparation of text with the help of auxiliary algorithms and the analysis itself carried out by the main algorithm. As an “input” for the auxiliary algorithms there is a real written text and as an “output” - the so-called hypothetical (output) text objects, which include hypothetical word forms (punctuation marks or sequence of letters, limited by spaces), hypothetical segments (sequences of word forms, limited by punctuation marks or conjunctions) and hypothetical phrases (sequences of segments, limited by a full stop). As an “input” for the main algorithms there is some prepared text, divided into hypothetical word forms, phrases and segments, and as an “output” - syntactic tree of dependencies or some other similar formal object, that explicitly represents the syntactic structure of each sentence. In some stage during the transition process from the beforehand prepared text to its syntactic tree, each hypothetical object becomes the ultimate or reduced. The final word form – is an analog of word or phraseological unit; the final segment - is an analog of a simple sentence (i.e., simple sentence, participial or other constructions), the final phrase equals to the full sentence [1, p. 234-241].

The main principles of automatic predictive analysis were outlined in the works of C. Hockett and then formalized by J. Rhodes and A. Oettinger. As an analysis tool of the syntactic structure of a sentence there is a set of syntactic predictions (SSP), which is stored in the computer's memory and contains a hypothetical list of possible syntactic structures in some sentence. The process of analysis consists in the sequential comparison of words of the analyzed sentence with all predictions, listed in the SSP. When the prediction is found for the word, then a mark of the corresponding syntactic function is attributed to it, and the necessary changes are made to SSP. The analysis continues for as long as every word form of the sentence gets the defined syntactic function [1, p. 242-243].

The fulcrum method is a further development of the ideas underlying the predictive analysis. The search of syntactic structure of a sentence begins with an identification of its fulcrums. The main sentence is considered as a fulcrum of a complex sentence and the predicate is considered as a fulcrum of a simple sentence. These fulcrums contain the maximum number of grammatical information and their detection can propose reasonable hypotheses (predictions) concerning other elements of the sentence. In order to conduct such an intended search of syntactic structure of the sentence, an automatic dictionary must be input in the machine memory. In this dictionary to each word a grammatical code is assigned, that represents the morphological properties of a word and points at all its potential syntactic features (government, agreement). Then syntactic search comes to definition of which possibilities inherent in the word are actually implemented in this sentence [1, p. 243-244].

The filter method is similar to the first two methods because its search of the syntactic structure of the sentence begins with the nomination and validation of some set of hypotheses, but its fundamental difference from the predictive analysis and fulcrum method is that at the beginning the set of hypotheses is not limited. The basic principles of filter method are formulated in the works of I. Leserf and developed by a number of other researchers, in particular, by L. N. Yordanskaya. These principles can be summarized as follows: the task is to find the syntactic structure of a text, i.e., to match each its sentence with some tree of dependencies. Identifying the tree of dependency starts with a review of the whole set of feasible solutions, each of which is a hypothesis about the possible syntactic functions of word forms of a sentence. Then the hypotheses are tested using a special program, and those which are no longer justified, are rejected. The program works, therefore, as a so-called filter that retains all the wrong decisions. The best, in some way, hypothesis is given to the output, and it is considered as the decision [1, p. 245].

In the above-mentioned approaches each sentence is compared with some predetermined benchmark. Thus, using syntactic models of analysis information about the syntactic structure of the sentences can be derived from text.

2.3. The semantic models of analysis

Modeling of an intelligent speech of human behavior was the subject of study to many scientists, including A. K. Zhovkovskyi, N. N. Leontyeva, Yu. S. Martemyanov, V. Yu. Rosenzweig, Yu. K. Shchekhlov et al. The semantic models of analysis simulate a person's ability to understand and create "semantically correct", i.e., meaningful sentences. In this case, not only the possession of grammar of a language is modeled, but also the possession of the meanings of words. The possession of the meanings of words is manifested in the ability of speaker to express the same opinion in different ways, and the ability of listener to understand the semantic identity of utterances which are different from outside. Thus, there is some semantic language or "language of thought" which is not given us under direct observation. Creation of the meaningful sentence can be represented as translation from semantic language into natural one, and understanding of the sentence - as translation from natural language into semantic one. A word's meaning in general is not an elementary semantic unit. It is divided into more elementary senses, which are the units of vocabulary of semantic language. A particular combination of elementary senses forms the meaning of some word of natural language. A small number of elementary senses gives a very large number of possible combinations of words of natural language. The units of semantic language – are those elements ("atoms"), with different combinations of which the "molecules" are made – i.e., the

meanings of the actual words of natural language. In general, an expression in the semantic language looks like a tree (or trees) of dependencies and, hence, the semantics can be described as the syntax [1, 252-255].

Thus, semantic language has its own words (elementary senses) and a grammar (trees of dependencies). In order to model a meaningful speech behavior during the perception of the text (analysis) one should have: 1) semantic dictionary of some language 2) algorithms of translation from the given language into the semantic one. In order to model a meaningful speech behavior during the creation of the text (synthesis) one should have: 1) semantic dictionary of some language 2) algorithms of translation from the semantic language into the given one [1, p. 262].

The algorithm of translation from the natural language into the semantic one is conducted in three stages. In the first stage, each part of the tree of syntactic dependencies, that contains a word that indicates the relations and all dependent on him words, is rewritten with the help of semantic dictionary as a tree of semantic dependencies. In the second stage, the semantic trees, similarly built for each part of syntactic tree, are combined according to certain rules in the semantic tree of the entire sentence. Thus, in the second stage, the complete sentence is fully translated into semantic language. In the third stage, the resulting record is reduced to a standard form using semantic equations. The simplest example of semantic equations is the rule of removing the double negation, which is equivalent to the statement [1, p. 263-264]. Thus, semantic models of analysis, with the help of predetermined information about grammar and meaning of words of the given language, model person's ability to understand and produce meaningful sentences.

Conclusions

In this paper the definitions of notions 'model' and 'modeling' in linguistics are reviewed. The steps of building a model and its basic properties are outlined. The main types of linguistic models and their application to solving a variety of linguistic tasks are described in detail.

Therefore, modeling of linguistic phenomena is still an extremely important today and should be further studied because there are still many unresolved issues.

1. Apresyan Yu. D. *Idei i metody sovremennoi strukturnoi linhvistiki (kratkij ocherk)*. – M.: Prosveshchenie, 1966. – 305 s. 2. Akhmanova O. S. *Slovar linhvisticheskikh terminov*. – 2-e izd., ster. – M.: Sov. entsiklopedia, 1969. – 608 s. 3. *Korpusna linhvistyka: Monogr.* / V. A. Shirokov, O. V. Buhakov, T. O. Hriaznukhina, O. M. Kostyshyn, M. Yu. Kryhin; NAN Ukrainy, Ukr. mov.-inform. fond. – K.: Dovira, 2005. – 472 s. 4. *Linhvisticheskii entsiklopedicheskii slovar / Hl. red. V. N. Yartseva*. – M.: Sov. entsiklopedia, 1990. – 683 s. 5. Losev A. F. *Vvedenie v obshchuiu teoriu yazykovykh modelei / Pod red. I. A. Vasilenko*. Izd. 3-e. – M.: Editorial URSS, 2010. – 296 s. 6. Revzin I.I. *Modeli yazyka*. – M.: AN SSSR, 1962. – 192 s. 7. Revzin I. I. *Sovremennaia strukturnaia linhvistika (Problemy i metody)*. – M.: Nauka, 1977. – 264 s. 8. *Sintaksicheski razmechennii korpus russkoho yazyka: instruktsiia polzovatelia*. [Electronic resource]. – Available from : — <http://ruscorpora.ru/instruction-syntax.html> 9. Bolshakov Igor, Gelbukh Alexander. *Computational Linguistics: Models, Resources, Applications*. Electronic edition IPN-UNAM-FCE, 2004. – 198 pp.