Linguistic Justification
for the Creation of a Lexical Analyzer
on a Semantic Basis

A. Hardzei

Minsk State Linguistic University
Minsk, Belarus
alieks2001@yahoo.com

Abstract—**The article examines in detail the logical, semantic and metasemantic paradoxes that arise when mixing theory and metatheory during the description of hierarchical systems, explains their significance for linguistics, provides examples of the internal contradictions in the Parts of Speech theory causing large-scale errors in the Natural Language Processing large-scale errors in the Natural Language Processing, and substantiates the creation of a lexical analyzer based on Parts of Language and the Theory for Automatic Generation of Knowledge Architecture (TAPAZ–2).**

Keywords—**paradox, Parts of Speech, Parts of Language, Signs of Syntax Alphabet, language categories, delineation procedures, lexical analyzer, Natural Language Processing, inventing machine**

Introduction

The reliability of syntactic rules depends entirely on the accuracy of determining the place of signs in the language hierarchy, i.e. on the correct interpretation for the content of linguistic patterns used by the intellect at a subconscious level as a basis for organizing the language by sequentially dividing it into parts. Any exceptions made here only indicate mathematical and semantic errors of the researcher, because language is determined in the same way as any other system. The correctness of the interpretation for the content of linguistic patterns, in turn, depends on the degree of its formalization – only then is it possible to prove the consistency of the consequences. In this regard, we recall the statement of A. Tarski that **the description of a language is only transparent when it leads to the creation of its formalized version** [1].

Why, despite the incredible efforts of scientists in various fields of knowledge, has natural language still not been formalized? Perhaps linguistic semantics cannot be formalized at all? Many linguists with a "classical philological education" are sure of this, although the founders of modern linguistics were not so categorical on this matter. Let us recall, at least, two striking, but opposite statements belonging to the famous European linguist Emile Benveniste:

"Some linguists reproach Saussure for a propensity to emphasize paradoxes in the functioning of language. But language is actually the most paradoxical thing in the world, and unfortunate are those who do not see this" [2, p. 37];

"To imagine a stage of language as "primal" as one would wish, but nevertheless real and "historical," in which a certain object would be *denominated* as being itself and at the same time something else, and in which the relation *expressed* would be a relation of permanent contradiction – a non-relating relation – in which everything would be itself and something else, hence neither itself nor the other, is to imagine a pure chimera" [2, p. 71–72].

Let us consider the paradox detected by B. Russell in 1902 in Foundations of Arithmetic by G. Frege, the first attempt at set-theoretic constructions. In any modern branch of knowledge, it is customary to operate with the concept of a set, which, as Yu. A. Shikhanovich wrote in his book Introduction to Modern Mathematics, "is so basic that it is difficult, at least today, to define it using simpler concepts" [3, p. 13]. **A set is the usual name for any collection of any objects – its elements**.

Most sets do not belong to themselves as elements. For example, the set of *cats* is not itself *a cat*, the set of *tables* is not itself *a table,* the set of *words* is not itself *a word*, etc. However, there are sets that belong to themselves as elements. For example, the set of *forests* is also itself *a forest*, the set of *sands* is also *sand*, the set of *waters* is also *water,* etc. Let us imagine some set *A*, the subsets of which include all sets that do not belong to themselves, for example, the set of all *hares*, which is not itself *a hare*, the set of all *boots*, which is not itself *a boot*, the set of all *letters*, which is not itself *a letter,* etc.

Now we consider the set *A* itself, recalling that it consists of all sets that do not belong to themselves. If the set *A* does not belong to itself, then it belongs to itself, because it includes all sets that do not belong to themselves; if it does belong to itself, then it does not belong to itself. **In any case, it turns out that the set *A* simultaneously belongs to itself and does not belong to itself** [4, p. 2].

Few linguists, even those who have read articles on formal semantics, know that this dramatic episode nearly drove the author of predicative second-order logic and founder of logical semantics to suicide, that G. Frege unsuccessfully tried to resolve this unfortunate paradox until the end of his life, and throughout the 20th century mathematics through the formalization of the meta-language had difficulty getting rid of the contradictions in its foundations. Neither predicate calculus [5], nor Boolean algebra [6], nor pseudo-physical and modal logic [7], nor Cantor's theory of sets [8] have been able to formalize the language semantics: **mathematics still did not have its own means for transformation of expressions and logic did not have its own means for its representation**. To the credit of the Belarusian science, essential progress in this area was achieved by V. V. Martynov in creating a semantic coding approach [9].

Contradiction in mathematics meant the collapse of bridges in engineering. The bridges of linguists did not fall until problems of the intellectual interface in human-computer communication arose, the main reason for which was, as V. V. Martynov repeatedly emphasized, **the ambiguity of statements in unlimited natural language** [9, p. 12].

The contradiction, which was called the mathematical paradox of B. Russell, turned into a disaster for the ideology of mathematics, causing a chain reaction of repeated proofs for theorems that seemed to have been solved once and for all. It still conceals the main danger for multidimensional mathematical constructions. Another, already semantic, paradox was detected by K. Grelling and L. Nelson in 1908 following Russell's paradox.

Let us group the adjectives so that the first ones – *autological* – denote a feature that they themselves possess, and the second ones – *heterological* – denote a feature that they themselves do not possess. For example, the adjectives *polysyllabic* (self also consists of many syllables) and *English* (self also belongs to the English language) are autological, in contrast to the heterological adjectives *monosyllabic* and *Russian*; but the adjective *русский* will already be autological.

Now we try to analyze the adjective *heterological* itself. If this adjective denotes a feature that it itself possesses, then it is non-heterological, and if it is non-heterological, then, by definition, it is heterological. Continuing the topic, we can cite the well-known paradox of the liar in its ancient or medieval interpretation: *"I am lying! – If at this moment I am really lying, then I am no longer lying, and if I am not lying, then, by asserting the opposite, I am lying"* (Eubulides); *"What Plato said is false,"* says Socrates. *"What Socrates said is true,"* says Plato* (Chrysippus) [10].

It is clear from the examples that **such contradictions arise when mixing the level and the meta-level, theory and meta-theory, language and meta-language**. The set *A* in Russell's paradox is a meta-level in relation to the subsets included in it, so considering the elements of the meta-level together with the elements of the level inevitably leads to a contradiction.

In other words, **paradoxes are a consequence of incorrect description of hierarchical systems, and not a property of these systems themselves**. Unfortunately or fortunately, almost all systems of our three-dimensional space are hierarchical, including natural language. Hence, the simplest and most reliable protection against paradoxes is to prohibit the inclusion of different level elements in the one level analysis, as well as the application of the same mathematical function to heterogeneous objects.

"To exclude such contradictions from an axiomatic theory, H. Rasiowa & R. Sikorski considered, it is necessary to describe its language precisely, i. e. the set of sentences of the theory and the set of signs used to build these sentences. In this way we avoid contradictions caused by a collision between the theory and its metatheory, that is, by including metatheoretical statements in the theory. This inclines us to introduce still greater precision in the construction of mathematical theories and leads to the concept of *formalized theories*, in which not only the properties of the primitive notions are given in an exact way by means of a set of axioms, but also the language of the theory is precisely defined" [11, p. 146–147].

However, for the formalized description of hierarchical systems, the greatest danger is posed by **metasemantic paradoxes** due to their hidden and fundamental nature. They affect the foundations of our primary perception of the World and manifest themselves in attempts to formally substantiate definitions and conclusions without taking into account their semantic counterparts. It is curious that semantics nevertheless seeps into formalisms in the form of **finite** or **non-finite technology**. It is no coincidence that the outstanding Italian mathematician G. Peano was inclined to the opinion that the set theory is more linguistics than mathematics.

As an example of inference rules, we will cite the paradox of the *conditional* or *material implication* *"if A, then B"*, which is true when: 1) its basis and its consequence are true; 2) the basis is false, and the consequence is true; 3) both the basis and the consequence are false, and establishing its truth does not presuppose a semantic connection between the statements. In classical logic, such an implication is false only in one case: when the basis is true, and the consequence is false; if *B* is true, then the entire implication is true, regardless of the truth of *A* or its semantic connection with *B*.

With this approach, the statements *"If there is life on the Sun, then Beijing is a big city"*, *"If two plus two is four, then it is time to plant potatoes"* are true. Conditional implication is true even when *A* is false, and

this does not take into account the truth of *B* or the connection of *A* with *B*. For example, the statement *"If the Sun is a cube, then the Earth is a triangle"* will be true. Obviously, **conditional implication does not fit well with a common sense understanding of cause and effect**.

As a result, logicians were forced to turn to semantics. In 1912, C. Lewis put forward the idea of ***strict implication*** [12], which partially reflected the semantic connection between the antecedent and the consequent in a conditional statement. However, it soon became clear that strict implication itself is not free of paradoxes, so in the middle of the 20th century, the German logician W. Ackermann [13] and the American logicians A. Anderson and N. Belnap [14] proposed replacing strict implication with ***relevant implication,*** which, in their opinion, resolved the paradoxes of both material and strict implication. Now implication was understood as logical consequence, i.e. the transition from premises to *reasonably* deducible conclusions, and the semantic definition of logical consequence took the form: *"From A follows B, if it cannot be that A is true and B is false"*.

Since the semantic connection between *A* and *B* in logical consequence **was declared but not formalized,** the meanings of the words "reasonably" and "in sense" could acquire different interpretations, so the relevant implication only postponed, but did not solve the problem of logical paradoxes. Moreover, the definition of logical consequence contained a hidden circle, since the words "is deducible" and "follows" used in it were synonyms of the word "implies". As shown by the cross-analysis of interpretations of the word meaning through a synonymous series conducted by Yu. N. Karaulov, the semantic circle occurs after a maximum of five occurrences [15].

An example of a semantic circle in the Dictionary of the Russian Language by S. I. Ozhegov: "To assure. The same as to convince (in 1 meaning)" [16, p. 713]. "To convince. 1. To assure" [16, p. 712]. As a result, the definition of the relevant implication automatically fell into the category of primary undefined concepts, thereby losing any semantic significance. With primary undefined concepts, the situation, from the metasemantic paradoxes point of view, was even worse.

It is known that in order to define some concept *A*, it is necessary to select a simpler concept *B* for it, and in order to define concept *B*, it is necessary to select an even simpler concept *C* for it; in the end, inevitably, there was some concept *D*, for which it was no longer possible to select a simple concept. In such a case, there was only one thing left: **to compare primary, undefined concepts with each other.** In other words, one can only talk about men in comparison with women. If nature had not created women, one would have to talk about some anthropomorphic creatures, but not about men.

The initial (primary, undefined) concept of the set theory, which became the basis of many sections of modern mathematics – general topology, general algebra, functional analysis, etc., is ***a set***, which is considered as a collection, a tuple, an ensemble of elements, etc. Here, the synonymous series has no meaning, since it inevitably leads to a semantic circle: **"a set is a collection, and a collection is a set."** Further, it is proposed to operate with "single-element" and even "empty" sets, which, from the standpoint of semantics, immediately levels the concept of "set": **if everything around is a set, including a single one, then there is no set**.

It is further proposed to consider that any set, in addition to proper subsets, has two "improper" ones – the set itself and some "empty" set, and there is an infinite number of such "empty" sets. Here, in addition to the hidden metasemantic paradox in the description of primary concepts, we have Russell's mathematical paradox, together with the paradoxes of Galileo [17], Hausdorff [18], and Banach-Tarski [19]. As we can see, **all the logical difficulties associated with the substantiation of the mathematical doctrine of infinity, when moving to the general theory of sets, only become even more acute**.

Since the second half of the 20th century, the influence of set theory on mathematics has significantly decreased due to the achievement of generalizations without explicit use of its apparatus, in particular, with the help of category-theoretical tools, by means of which it was possible to generalize practically all variants of set theory in *topos* theory. American mathematicians F. Lawvere [20] and M. Tierney [21] defined categorical analogues of operations on sets using **the basic categorical construction of the limit – an elementary topos,** i. e. a Cartesian closed category with a classifier of sub-objects. In a similar way, as it turned out, it is possible to categorically describe all possible constructions for the formation of new sets from existing ones.

Intuition convinces us that the infinity of combinatorics generated by a finite alphabet and a finite set of rules for construction and transformation has some limitations in hierarchical systems. Why, for example, do Russians say *"достать из-под шкафа" ("to get something from under the cupboard")* but not *"достать из-над шкафа" ("to get something from above the cupboard")*, i. e. from the mezzanine? Why do they say *"выйти в коридор" ("to go out into the corridor")* but not *"войти в коридор" ("to enter the corridor")*? Apparently, such limitations are purely semantic. The ancient Greeks and Romans felt this well in the example of the paradox of Hercules and the tortoise: the infinity of small numbers was somehow miraculously resolved. In other words, **infinity arises where we do not understand something, and operating with it without semantics leads to scholasticism**.

If we want to build a working system for artificial

modeling of intellectual activity, then the design of its knowledge base should be homomorphic to the architecture of the World Model created by the intellect in a natural way, i. e. it should meet the principles of visibility, obviousness and compliance with common sense. Of the existing scientific theories, the most suitable for this are **the Topos Theory, the Theory of Groups, the Euclidean-Hilbert Geometry [22], Quantum Physics and the Universal Semantic Code by V. V. Martynov**.

"At this point the problem that haunts all of modern linguistics arises: the relationship between form and meaning, wrote Emile Benveniste. Many linguists would like to reduce it to the notion of form alone but somehow they cannot succeed in freeing themselves of the correlative, meaning. What has not been attempted in order to avoid, ignore, or expel meaning? It has been useless; this Medusa's head is always there at the center of language, fascinating those who contemplate it" [2, p. 106–107].

I. Contradictions in the foundations of the Russian language academic grammar

Let us check some popular linguistic theories for paradoxes. Russian grammarians suggest analyzing sentences of the form *"Отец – учитель"* (literal translation *"Father – a teacher"*) according to the scheme GS(N_1) – GP(N_1): "Grammatical Subject is a name in the nominative case: grammatical predicate is a formally subordinate form of the name; in the paradigm structure, the nominative case in the grammatical predicate can alternate with the instrumental" [23, p. 551]. This approach does not stand up to criticism even at the first reading, since the verb *быть* (*to be*) appears in the past or future tense: *"Отец был (будет) учителем"* (*"Father was (will be) a teacher"*), and even in the present tense the verb is not difficult to restore: *"Отец есть учитель"* (*"Father is a teacher"*).

"It emerges, first, that the nominal sentence cannot be considered as lacking a verb. It is just as complete as any other verbal utterance. Furthermore, we cannot consider it a sentence with a zero copula, for in Indo-European there are no grounds for establishing a zero-form/full-form correspondence between the nominal sentence and the verbal sentence containing "to be". <. . .> Otherwise, there is not really an equation between the subject and the nominal term with verbal function" [2, p. 137–138] [1].

It is curious that the authors of Russian grammars are not even embarrassed by the fact that the role of the grammatical predicate is played by a noun in the nominative case, although traditionally such a form of the noun is assigned to the grammatical subject. As a result, it is necessary to resort to tricks, calling the phrase *"был*

(будет) учителем" (*"was (will be) a teacher")* the grammatical predicate and qualifying it as a compound nominal. Here, there is already a mixing of levels in syntactic analysis – the role of the grammatical subject is played by a word, but the role of the grammatical predicate is played by a group of words (a words-combination).

The relationship between a word and a combination of words is hierarchical: a word is included in a combination of words, a combination of words is included in a sentence, a sentence is included in a text. The paradox is manifested in the impossibility to continue analyzing the sentence when building up its syntactic structure *"Father was (will be) a very good teacher of Russian"*, since the entire part of the sentence after the word *"father"* automatically becomes the grammatical predicate. If syntactic analysis were accompanied by at least minimal semantic support, relied at least on school definitions of the grammatical subject and grammatical predicate, because, as H. Reichenbach wrote, "pseudo-problems arise, if we look for truth where definitions are needed" [24, p. 15], it would be possible able to avoid a dead end in parsing the sentence. Who is the sentence talking about? – About *the father "father" is the grammatical subject.* Who is (was / will be) the father? – Is (was / will be) *a teacher.* Or: What properties does (did, will) the father have? – The properties of *a teacher "are (were / will be)"* is the grammatical predicate and *"a teacher"* is the grammatical object. Next step, the grammatical attributes *"good"* and *"language"* are detected for the grammatical object *"teacher"* and then the grammatical attributes *"very"* and *"Russian"* are detected for the grammatical attributes *"good"* and *"language"*. The syntactic formula of the sentence is: $GS + [GP] + Attr_2 + Attr_1 + GO + Attr_4 + Attr_3$, where GS is the grammatical subject, GP is the grammatical predicate, GO is the grammatical object, Attr are the grammatical attributes with indices of the property accumulation order; the + (plus) sign indicates the order by which parts of the sentence are connected, and the brackets *[]* indicate a possible reduction of the grammatical predicate under certain conditions.

Two conclusions follow from the analysis. First: in Russian, **the role of a grammatical object can be played by a sign in the nominative case,** which is easy to agree with, since in grammatical objects the forms of the nominative and accusative cases tend to coincide, sometimes despite the animate or inanimate nature of the nouns.

"The category of inanimate nouns includes nouns denoting a collection of living beings *(народ "people", толпа "crowd", взвод "platoon", стая "flock", рой "swarm", группа "group"*, etc.), as well as collective nouns such as *молодёжь "youth", крестьянство "peasantry", детвора "children", пролетариат "proletariat"*, etc. <. . .> For animate nouns of the neuter

---

[1] Compare: *"A father is a teacher, but a teacher is not necessarily a father (a teacher maybe a mother and so on)"*. For more details, see: Hardzei A. The Principles of Calculating the Semantics of Subject Domains. Minsk, Belarusian State University Publ., 1998, pp. 33–35. (In Russian).

gender, as well as for inanimate ones, the accusative case form in the singular coincides with the nominative case form. For example: *Ax, как я люблю это пустое существо! – простонал Павел Петрович "Oh, how I love this empty creature! -– moaned Pavel Petrovich"* (Turgenev). The same is observed in feminine nouns with a zero ending in the nominative case: *вижу рысь, мышь "I see a lynx, a mouse"*. <...> The names of microorganisms fluctuate between animate and inanimate nouns: *микроб "microbe", бацилла "bacillus", инфузория "infusoria", бактерия "bacterium", амёба "amoeba"*, etc. They have two accusative case forms: *изучать микробов* and *микробы "to study microbes"; рассматривать вирусов* and *вирусы в микроскоп "to examine viruses under a microscope"; уничтожать бацилл* and *бациллы "to destroy bacilli"*. <...> The words *тип "type", образ "image", характер "character"* are used as inanimate nouns, which are the names of characters in works of art: *создать сильный характер "to create a strong character"; охарактеризовать отрицательные типы и положительные образы; "to characterize negative types and positive images""* [25, p. 100–101].

Second: the comparison degrees of adjectives and adverbs need to be clarified, since the words *очень (сильно) "very (much)", самый "most", чрезвычайно "extremely"*, etc. not only play an independent role in the sentence, allowing the question *очень какой? – очень хороший "very what? -– very good", самый какой? – самый красивый, "most what? – most beautiful", чрезвычайно какой? – чрезвычайно быстрый "extremely what? – extremely fast"*, but are also able to act as an independent sentence: *Ты хочешь пойти в театр? – Очень! "Do you want to go to the theater? – Very much!"*, therefore they cannot be considered auxiliary means of constructing analytical forms of superlative degrees of adjectives or adverbs, which means that these degrees do not exist, otherwise, we again get a mixture of the level (word) and the meta-level (words combination) and face Russell's paradox.

Incidentally, it should be noted that equating the combination of a functional sign with a notional sign to an ordinary fragment of phrase, i. e. to a combination of notional signs, is a serious mistake, since the so-called "functional Parts of Speech" do not belong to language, but to metalanguage, and denote facts of language, not facts of the World Model. For example, the preposition *на "on"* is used to construct locatives of the form *на столе "on the table"*, but not to denote individs that play the role of a locus in the World Model. The preposition *на "on"* disappears if the structure of the sentence is brought into line with the structure of the World Model: "The surface of the table holds the book" instead of "The book lies on the table". Since "functional Parts of Speech" do not play independent roles in a sentence and are not parts of the sentence, in order to avoid

Russell's paradox, they must be excluded from the Parts of Speech and classified as **Signs of Syntax Alphabet, i. e. auxiliary syntactic means (at the macrolevel – prepositions, postpositions, particles, conjunctions, etc., at the microlevel – flexions, prefixes, infixes, postfixes, etc.) that serve for connecting the components of language structures and the formation of morphological paradigms**.[2]

Let us consider the sentence *Он очень быстро бегает "He runs very fast"*. It is obvious that *он "he"* and *бегает* "runs" are related as an individ and his temporary characteristics. Compare: *"If a doctor in the forest chops down a tree, then at this time he is not a doctor, but a woodcutter"*. Chopping down trees is a temporary occupation for a doctor. If the doctor refuses to heal and only chops down trees, then he really does turn into a woodcutter, which is confirmed in the language by semantic-syntactic transformations: *врачевать людей "heal people" > врачующий людей "(a man that) healing people" > врач "healer"; рубить лес "cut the wood" > рубящий лес "(a man that) cutting wood" > лесоруб "woodcutter"*. Let us denote the individual *он "he"* by the symbol i, then the process *бегать "to run"* will be written as $f(i)$. If *бегает "runs"* is characteristics of the individ *он "he"*, then *быстро "quickly"* is characteristics of his characteristics. It is impossible to say *он быстро "he quickly"*, we can only say that *быстро бегает "runs quickly"*. In other words, the relations between the characteristics *быстро "quickly"* and *бегает "runs"* are hierarchical, which indicates a superposition of the functions $f(f(i))$ or the degree of the function $f^2(i) = f(f(i))$. If *бегает "runs"* is characteristics of the individ *он "he"*, and *быстро "quickly"* is characteristics of his characteristics, then *очень "very"* is characteristics of his second characteristics, i. e. characteristics of characteristics of his characteristics. We can not say *он очень "he very"*, just as we can not say *очень бегает "runs very"*, we can only say *очень быстро "very quickly"*, so the relationships between the characteristics *очень "very"* and *быстро "quickly"* are also hierarchical: $f^3(i) = f(f^2(i)) = f(f(f(i)))$. Thus, in the hierarchy of characteristics *бегает "runs"* has the first degree, *быстро "quickly"* has the second degree, and *очень "very"* has the third degree. The adverb in Russian comes from the short form of the neuter adjective: *он хороший → он хорош "he is good", она хороша "she is good", оно хорошо "it is good"*.[3] It is not surprising that the paradigms of their degrees are close and intersect, to the point of erasing the boundary

[2] The term *Signs of Syntax Alphabet* is borrowed from: Rasiowa H., Sikorski R. The Mathematics of Metamathematics. 2nd edn. Waszawa, Polish Scientific Publishers, 1963, pp. 151–154.

[3] On the primacy of the full and secondary nature of the short form of the Russian adjective, see: Kurilovich E. Lexical Derivation and Syntactic Derivation. In : Essays on Linguistics. Moscow, Foreign Literature, 1962, p. 60 [26].

in intensive forms: *самый хороший* "the best", but *очень хорошо* and *очень хороший* "very good"; *строгий* → *строже* → *более строгий* and *строго* → *строже* → *более строго* "strict → stricter → *more strict*". By the way, the intensive form allows only the question *how?* – *строже* "stricter", but not at all *which one?* The hierarchy of degrees of constant characteristics can be conveniently traced using the example of *самый красивый бежевый цвет* "the most beautiful beige color". What color is i? – It is *бежевый* "beige" $p(i)$, what kind of *бежевый* "beige"? It is *красивый* "beautiful" $p(p(i)) = p^2(i)$, what kind of *красивый* "beautiful"? It is *самый* "the most" $p(p(p(i))) = p(p^2(i)) = p^3(i)$. Although in Russian we can say *красивый цвет* "a beautiful color" (when **самый цвет* "the most color" or **самый бежевый* "the most beige" is inadmissible), nevertheless, *красивый* "beautiful" occupies the second position in the hierarchy of characteristics, because in an unclear situation, the restoration of the first characteristics is required. What color exactly is characterized as *beautiful*? – It is *beige*. It follows that *бежевый* "beige" has the first degree, *красивый* "beautiful" has the second degree, and *самый* "the most" has the third degree. It is generally accepted that degrees of comparison are the prerogative of the majority of qualitative adjectives [25, p. 129]. It is not explained why not all of them. But with the correct semantic analysis, the answer is simple: **intensive evaluative forms are absent from all signs of the first degree,** regardless of their "quality". These include, for example, signs of the basic color palette: *белый* "white", *жёлтый* "yellow", *красный* "red", *зелёный* "green", *синий* "blue", *чёрный* "black", etc., if they are not to be confused with homonymous epithets: *белый* "white" – *ясный* "clear", *чистый* "pure", *бледный* "pale"; *жёлтый* – *старый* "old", *больной* "sick", *увядший* "withered"; *красный* "red" – *горячий* "hot", *красивый* "beautiful", *стыдливый* "bashful"; *зелёный* "green" – *незрелый* "immature", *молодой* "young", *неопытный* "inexperienced"; *синий* "blue" – *уставший* "tired", *замёрзший* "frozen", *побитый* "battered"; *чёрный* "black" – *тёмный* "dark", *грязный* "dirty", *мрачный* "gloomy". Metaphors are conditioned by the characteristic color symptoms of the properties manifestation. Compare: *чёрный* "black" → *чернее* "blacker" = *мрачнее* "darker", *но коричневый* "brown" → *(?)*, *оранжевый* "orange" → *(?)*, *ультрамариновый* "ultramarine" → *(?)*. The division of adjectives into qualitative, relative and possessive also raises doubts. Firstly, opposition is always binary. A ternary opposition arises when the whole is first divided into two parts, then one part is divided into two more, with the other part remaining undivided, which results in a confusion of the level and the micro level, and we face Russell's paradox. Secondly, in epistemology it is customary to contrast *quality* with *quantity*. But

the relative adjectives *деревянный* "wooden", *утренний* "morning" and the possessives *мамин* "mother's", "*учительский*" teacher's denote the same quality as the "qualitative" *круглый* "round" and *мятый* "crumpled", as opposed to the adjective-numeral *второй* "second" (which / what?). Grammarians of the Russian language claim that qualitative adjectives directly name various attributes of objects, while relative ones do so indirectly: "The attribute of an object in them indicates various relationships: to an object (compare: *железная кровать* "iron bed" vs. *кровать из железа…* "a bed made of iron …"), to a person (compare: *мамино пальто* "mother's coat" vs. *пальто мамы…* "coat belonging to mother..."), to a place (compare: *загородная поездка* "a town out trip" vs. *поездка за город…* "a trip out of town …"), to time (compare: *утренние сообщения* "morning messages" vs. *сообщения, сделанные утром* "messages made in the morning …"), to an action as a property (compare: … *раздвижной стол* "apart sliding table" vs. *стол, который раздвигается* "a table that slides apart…"), to a number (compare: *шестой стол* "sixth table" vs. *шесть столов* "six tables…"), etc." [25, pp. 127–128]. However, upon closer examination it turns out that qualitative adjectives also indicate a variety of relationships: *sweet, bitter* – to taste; *white, blue* – to color; *sonorous, quiet* – to sound; *soft, smooth* – to coating; *narrow, long* – to shape; *wise, kind* – to individual character, etc. It would be correct to divide adjectives first into qualitative and quantitative, and then qualitative ones into relative and possessive. Then it would be clear that in the combinations of *отцов* ремень "fathers' belt", *мамина* сумка "mother's bag", *собачья* будка "doghouse", *кофейный* запах "coffee smell", and finally, *лисья* морда "fox's muzzle", the role of definition is played by possessive adjectives, and in the combinations of *кофейный* сервиз "coffee set", *лисья* шуба "fox's fur coat", by relative adjectives, because <u>a fox's muzzle</u> belongs to a fox, unless it is a metaphor *(Ах ты, лисья морда! "You, fox's muzzle!")*, and <u>a fox's fur coat</u> no longer belongs to a fox, unless it is a metonymy *(По сугробам бегала лисья шуба "A fox's fur coat ran through the snowdrifts")*.

Unfortunately, the contradictions in the classification of Speech Parts are fundamental in nature and it is impossible to eliminate them all. For example, along with nouns, adjectives, verbs and adverbs denoting, respectively, objects, attributes of objects, processes and attributes of processes, pronouns and numerals are distinguished as independent Parts of Speech, which in content are varieties of nouns *он* "he", *это* "this (is)", *восемь* "eight" or adjectives *его* "his", *этот* "this (thing)", *восьмой* "eighth". The word *столовая* "canteen" (place for eating), contrary to morphology, is considered a noun, while the word *бег* "running" with verbal semantics is also among the nouns. Interjections include words

expressing but not naming emotions and expressions of will [25, p. 257]. At the same time, in the phrase *А девица xu-xu-xu да ха-ха-ха! "And the maiden hi-hi-hi and ha-ha-ha!"* the words *xu-xu-xu "hi-hi-hi"* and *xa-xa-xa "ha-ha-ha"* are considered interjections, although they denote a process and play the role of a grammatical predicate [27].

The listed errors are sufficient to agree not only with the restrained statements of F. de Saussure and O. Jespersen that "We are then dealing with a defective or incomplete classification; the division of words into substantives, verbs, adjectives, etc. is not an undeniable linguistic reality" [28, p. 109–110], and their "definitions are very far from having attained the degree of exactitude found in Euclidean geometry" [29, p. 58], but also with the rather harsh statements of L. Tesnière: "This classification, which rests on vague and sterile **empiricism** and not on a precise and fertile doctrine, does not stand up to scrutiny. <...> All evidence indicates that a good classification cannot rest **simultaneously** on multiple characteristics. We must therefore separate essential characteristics from secondary ones, or, to borrow terminology from logic, to distinguish dominant characteristics from subordinate ones. The principle of subordination leads to the establishment of a hierarchy of criteria, the most important of which serves to determine the **classes** and the second most important of which serves to determine the **subclasses**" [30, p. 45] and F. F. Fortunatov: "The division into Parts of Speech that is accepted in our grammars (and has come to us from ancient grammarians) represents a mixture of grammatical classes of words with their non-grammatical classes and therefore cannot have scientific significance (emphasis added – *A.H.*). For example, 1) verb, 2) noun, 3) pronoun in Russian, Greek and Latin do not represent correlative classes of words in the grammatical classification of words..., but in the non-grammatical classification they also do not form correlative classes of words, since, for example, the difference between a noun and a pronoun on a non-grammatical basis has nothing in common with the difference between a noun and a verb" [31, p. 166].

It is impossible to formalize a contradictory doctrine in which any statement is "provable". It is not surprising that the developers of *the Semantic Web* project, despite their titanic efforts to standardize the technology, have still not succeeded in achieving machine readability of Internet pages in the direction from morphological analysis to syntactic and semantic, as well as constructing a top-level ontology "from below", i. e. by reducing to it numerous empirical subject ontologies implemented in different standards [32], since **there is no one-to-one correspondence between morphology, syntax and semantics**. "Language is a process of communication between people, and is inextricably enmeshed in the knowledge that those people have about the world. That knowledge is not a neat collection of definitions and axioms, complete, concise and consistent. Rather it is a collection of concepts designed to manipulate ideas. It is in fact incomplete, highly redundant, and often inconsistent" [33, p. 26]. If the goal of Web semantization set by the head of the W3C Consortium T. Berners-Lee in 2001 [34] had been achieved, then modern computing power, both hardware and software, including large language models and other neural network algorithms for deep machine learning and big data processing, would already have allowed the Internet to be transformed into a Global Artificial Intelligence – so called *Giant Global Graph (GGG)*. However, first it will be necessary to solve the problem of automatic semantic markup of natural language content, the main stumbling block of which was **the incorrect operation of the lexical analyzer, which generated critical errors in syntactic analysis and catastrophic errors in semantic analysis**.

II. Language categories

Metasemantics of language categories or semantics of **language semantic categories** is the content of language patterns that form the basis for distinguishing parts of language in a paradigm [35]. The reliability of syntactic rules depends entirely on the accuracy of determining the place of signs in the language hierarchy, i. e. on the correct interpretation of the language patterns content used by the intellect at a subconscious level as the basis for ordering the language by sequentially dividing it into parts. Any exceptions made here only indicate mathematical and semantic errors on the part of the researcher, because language, like any other system [4], is always determined [5]. The correctness for the interpretation of a linguistic pattern content, in turn, depends on the degree of its formalization – only then is it possible to prove the consistency of the consequences. The Theory for Automatic Generation of Knowledge Architecture (TAPAZ), which includes the Deductive Theory of Language and its Pre-Description [6], is effective in solving this issue and allows us to identify the following semantic categories.

A. The World Model, language and metalanguage

The primary level of signs semantic stratification consists of **Language Parts** – subsets of the language

[4] See the works on general systems theory by A. A. Bogdanov [36], T. Kotarbinski [37], L. von Bertalanffy [38], K. Boulding [39], N. Wiener [40], W. R. Ashby [41], J. von Neumann [42], etc.

[5] Contradictions arise not in the system, but during its implementation in a material where some system has already been implemented, and therefore are not systemic, but intersystemic in nature. Hence, for example, the *ideal pendulum* system, widely known in physics and described by differential equations, is easier to implement in metal or stone than in wood or paper. For more details on this, see: Melnikov G. P. Systemology and Language Aspects of Cybernetics. Moscow, Sov. Radio, 1978, 368 pp. (In Russian).

[6] See: Hardzei A. The Principles of Calculating the Semantics of Subject Domains. Minsk, Belarusian State University Publ., 1998, 156 pp. (In Russian).

system, the elements of which are signs with a common, extremely abstract pattern in their aspect of meaning, and **Signs of Syntax Alphabet** – auxiliary means of syntax (at the macro level – prepositions, postpositions, conjunctions, particles, etc., at the micro level – inflections, prefixes, postfixes, infixes, etc.), serving to connect the constituents of language structures and the formation of morphological paradigms [43, p. 11]. Signs of Syntax Alphabet do not belong to the language, but to the metalanguage (the language that interprets another language), because they denote not the facts of the World Model, but the facts of the language; they do not play independent roles in a sentence.

B. The individ and individ's attribute [7]

Parts of Language are subsets of the language system, the elements of which are signs with a common, extremely abstract pattern in their aspect of meaning. They are divided into *substantives (taigens)*, denoting individs, for example: *книга "book"*, *стол "table"*, *восемь "eight"*, *мы "we"*, and *predicatives (yogens)*, denoting the attributes of individs, for example: *бежать "run"*, *коричневый "brown"*, *смело "boldly"*, *очень "very"*. They differ in four parameters:

1) *semantic* – if, as a result of substituting the diagnosed element to the left of … *causes* …, a marked sentence is obtained, then it is a predicative, if unmarked – then a substantive, for example: *Бег вызывает усталость "Running causes fatigue"*, but *"*Does the city cause …?"* [44, pp. 17–19]; to increase the reliability of the test and to exclude metaphors like **Город вызывает…? "The city causes (evokes) admiration"*, we recommend using the opposite procedure: if to the right of …*, when* … the diagnosed element in the role of the grammatical predicate turns the sentence into a correct one and semantically identical to the sentence with … *causes* …, then it is a predicative, otherwise it is a substantive, and any rewriting of affixes is allowed, for example: *Я устаю тогда, когда бегаю "I get tired, when I run"*, but in the case of **Я восхищаюсь тогда, когда горожу? "*I feel admiration, when I city?"* the correct sentence is *Я восхищаюсь тогда, когда нахожусь в городе "I feel admiration, when I am in the city"* with the diagnosed element in the role of the grammatical circumstance, therefore, *бег "run"* is a predicative, and *город "city"* is a substantive [45, p. 71];

2) *syntactic* – in an expanded sentence, predicatives occupy central positions, substantives – marginal ones [44, pp. 17–25], for example: *"Спутник меняет орбиту" "The satellite changes orbit"*;

3) *syntagmatic* – in non-composite taigens the modificator (defining component) comes first, and the actualizator (defined component) comes second, which is often collapsed into a suffix or erased: *принимающее **устройство** "receiving**device**"* → *приём-**ник** "receiv-**er**", булочная **лавка** "bakery**shop"** → булочная-**ø** "bakery-**ø**"*; in non-composite yogens, on the contrary, the actualizator comes first (often collapsed into a prefix or erased: ***делать** круг "**make** a circle"* → *о-кружать "en-circle", **бежать** галопом "**run** at a gallop"* → *ø-галопировать "ø-gallop"*), and the modifier comes second [8];

4) *paradigmatic* – predicatives have a degree, substantives do not. Variable predicatives of the *first degree* denote a variable attribute of an individ, that is, a process in which he participates (e. g., *to squeeze*); variable predicatives of the *second degree* denote an attribute of an attribute of an individ (e. g., *quickly* → *to run quickly*) [9]; variable predicatives of the *third degree* denote an attribute of an attribute of an attribute of an individ (e. g., *very* → *quickly* → *to run*). The second and third degrees of predicatives can be positive or negative. Predicatives of the *second* and *third positive degrees* clarify the meanings of predicatives of the first degree (e. g., *to run very quickly*), and predicatives of the first degree – the meanings of predicatives of the second and third negative degrees (e. g., *very much* → *to love* ← *to run,* i. e. *to love running very much* [10]). Positive or negative sign of the third degree in predicatives is determined syntagmatically: if a predicative of the third degree relates to a predicative of the second positive degree *(to run very quickly)*, then it has a positive sign, if to a predicative of the second negative degree *(to love reading very much)*, then

---

[7] For two types of pattern, see: Hardzei A. Foundations of Combinatory Semantics. L. V. Rychkova (ed.) Collection of Papers on Lexicography Vocabulum et vocabularium [The Word and the Lexicon]. Grodno, Grodno State University Publ., 2005, pp. 32–35. (In Russian); Hardzei A. Semantics of Metasemantics. In : Scientific Notes of the V. I. Vernadsky Tauride National University. Series : Philology, vol. 20 (59). 1, Simferopol, 2007, pp. 126–133.

[8] On the nominative unit and its contraction, see also: Hardzei A. The Principles of Calculating the Semantics of Subject Domains. Minsk, Belarusian State University Publ., 1998, 156 pp. (In Russian); Martynov V. V. Language Categories. The Semiological Aspect. Moscow, Science Publ., 1982, 192 pp. (In Russian); Martynov V. V. Semeiological Foundations of Computer Science. Minsk, Science and Technology Publ., 1974, 192 pp. (In Russian); Martynov V. V. Foundations of Semantic Coding. Experience of Knowledge Representation and Transformation. Minsk, European Humanitarian University Publ., 2001, 140 pp. (In Russian); Rozwadowski, J. Słowotwòrstwo i znaczenie wyrazòw. In : Wybòr pism. Warszawa, PWN, 1960, t. 3. Językoznawstwo ogòlne, s. 21–95. (In Polish).

[9] The name *evaluative* is appropriate for variable predicatives of the second degree, since their semantics can be measured: *быстро "fast"*, *быстрее "faster"*, *очень быстро "very fast"*.

[10] Predicatives of the second negative degree are conveniently called **modal** by analogy with modal verbs in the Greco-Latin classification.

it has a negative sign [45, pp. 74–75].

We note that the semantic categories of *the World Model, individ,* and *individ's attribute* pertain to *the World Theory*; *language, Part of Language, taigen (substantive), yogen (predicative)* pertain to *the Theory of Language*; *metalanguage, Sign of Syntax Alphabet* pertain to *the Theory of Metalanguage*. In our further presentation, the names of the categories of the World Theory and the Language Theory will coincide (for example, *single-place* vs. *multi-place* characterize both the World Model and the Language), but they should be distinguished, since the semantic categories of the World Theory depend on the observer's point of view (at the micro level, *a tree* is perceived as a system of individs *root + trunk + crown,* at the macro level – as a whole individ, *a separate tree*), and the semantic categories of the Language Theory and the Metalanguage Theory, i. e. linguistic and metalinguistic semantic categories are independent (the Russian sign *дерево "a tree"* always denotes one individ, *деревья "trees"* denote a group of individs; the Russian preposition *в "in"* is always used to construct locatives of inclusion *положить книгу в ящик письменного стола "to put a book in a desk drawer",* but not exclusion **положить книгу вне ящика письменного стола "*to put a book outside a desk drawer"*). A strict distinction between the categories of the World Theory, the Language Theory, and the Metalanguage Theory is necessary to achieve consistency of analysis and avoid Russell's paradox.

C. Expanded and reduced formation

In developing the idea of A. V. Isachenko on semantic condensation [46, p. 338], Parts of Language with minimal semantic condensation are called **expanded**; these are either *composite* stable combinations of signs with an obvious two-component structure of free (morphologically integrally formed) stems and/or roots, such as *хвойный лес "coniferous forest"*, *научно обоснованный "scientifically proven"*, *покрыть краской "cover with paint"* or *complex* signs with a two-component structure of connected (truncated) stems and/or roots, such as *водопад "waterfall"*, *солнцезащитный "sunscreen"*, *злословить "backbite"*. **Reduced** Parts of Language are those with the maximum degree of semantic condensation, these are either *contracted* Parts of Language with hidden two-component combinations of connected (truncated) stems or roots with stems or roots condensed into affixes, such as *приём-ник "receiv-er"* ← *принимающее устройство "receiving device"*, *пре-красный "beauti-ful"* ← *очень красивый "very beauty"*, *о-кружать "en-circle"* ← *делать круг "make a circle"*, or *constricted* Parts of Language, which are combinations of free stems or roots with erased but semantically

reconstructed stems or roots [11], such as *лазер "laser"* ← *light amplification by stimulated emission of radiation*, *булочная-ф "bakery-ф"* ← *булочная лавка "bakery shop"*, *ф-галопировать "ф-gallop"* ← *бежать галопом "run at a gallop"*. The ambiguity of categorical semantics of taigens and yogens in diachrony creates favorable conditions for their conversion in synchrony by rearranging taigens from marginal positions of a syntactically expanded sentence to central ones, and yogens from central to marginal ones: *(Я) ремонтирую дверной замок "I repair the door lock"* vs. *(Я) замыкаю дверь "I lock the door"*. In synchrony, conversion closely interacts with affixation and in the process of sign compression can be considered as its degenerate case, especially in languages with developed morphology: *столовая комната "canteen area"* → *столов-ка "cant-ina"* → *столовая-ф "canteen"*.

D. Informational and physical localization

A fragment of the World Model, including the roles of all participants in the event [12], is defined as physical or informational by the nature of the interaction of individs: if the subject's shell acts as an instrument, then the impact is physical: *to treat* (physical yogen) → *doctor, medicine, patient* (physical taigens), if its surroundings, then it is informational: *to teach* (information yogen) → *teacher, knowledge, student* (information taigens).

E. Constant and variable feature

Substantives (taigens) and predicatives (yogens) are divided into **constants** and **variables,** depending on whether they denote the sets of homogeneous individs *i-const,* for example: *'idea', 'horse' (constant substantives)* or the sets of heterogeneous individs *i-var,* for example: *'it', 'this' (variable substantives)*, whether they denote the set of properties of individs $p(i)$: *'mental', 'gray' (constant predicative)* or the sets of functions of individs $f(i)$, i. e., the processes in which individs are involved: *'to think', 'to gallop' (variable predicative)*. Some syntactic rules, for example, the order of grammatical attributes, require additional division of constant taigens and yogens into **proper** and **common**, which select an individ and individ's attribute from a homogeneous set *Минск (city Minsk)*, *минский (Minsk region)*, or, conversely, include in a homogeneous set *город "city"*, *городской "urban"*; besides, it is advisable to divide variable taigens and

---

[11] For etymological procedures for reconstructing the two-component nature of signs, see: Martynov V. V. Etymology and Hidden Two-Component of the Word. In : Studia etymologica Brunensia 2. Praha, 2003, pp. 81–91. (In Russian).

[12] **Subject** – the originator of the action, **object** – the recipient of the action, **instrument** – a performer of the action, the closest individ to the subject, **mediator** – i. e. mediator of the action – the closest individ to the object. For more information on the World Model and its structure, see: Hardzei A. Foundations of Combinatory Semantics. L. V. Rychkova (ed.) Collection of Papers on Lexicography Vocabulum et vocabularium [The Word and the Lexicon]. Grodno, Grodno State University Publ., 2005, pp. 32–35.

constant yogens condensed from them into **personal** and **impersonal**, denoting the first person (persons) in the role of an observer (observers) *'I', 'my' ('we', 'our')*, the second person (persons) in closest environs of the observer (observers) *'you', 'yours' ('you', 'yours')*, and the third person (persons) in distant environs of the observer (observers) *'he', 'she', 'it', 'his', 'hers' ('they', 'theirs')* or, conversely, not directly denoting anybody in environs of the observer (observers) *'someone', 'everyone'* [13].

F. Qualities and quantities

The procedure of delimitation is based on the passage to the limit, the most important operation of mathematical analysis. "The basis of this operation is the fact that the distance between any two points on the real line is defined. A number of fundamental facts from analysis are not connected with the algebraic nature of the set of real numbers < ... >, but depend only on those properties of real numbers which are related to the concept of distance" [47, p. 16]. If a sign denotes a number or a feature (property, function) of a number or a feature (property, function) of number and a geometric figure, then it is quantitative; if a sign denotes a geometric figure or a feature (property, function) of a geometric figure, then it is qualitative, for example, *два "two"*, *второй "second"*, *складывать "to add"* are quantitative signs (constant taigen, constant yogen and variable yogen), *дуэт "duet"*, *двустворчатый "bivalve"*, *соединять "to connect"* are qualitative. It is convenient to call quantitative constant informational taigens and yogens **numerical**, [14] quantitative variable informational and physical taigens – **measured.** [15] Qualitative variable taigens and qualitative constant reduced yogens are additionally divided into **indicative** *это "this (is)"*, *то "that (is)"*, *этот "this (thing)"*, *тот "that (thing)"*, and **arbitrary** *белый "white"*, *каменный "stone (house)"*, indicating and, accordingly, not indicating an individ and individ's attribute in the closest or distant environs of the observer. Numerical and arbitrary taigens are divided into **interrogative** *сколько "how many"*, *кто "who"* and **narrative** *один "one"*, *все "all"*, [16] and arbitrary yogens are divided first into **relative** *белый "white"*, *картонный "cardboard (packaging)"*, denoting features that relate to an individ, and **possessive** *чей "whose"*, *собачий*

"dog's", denoting features that belong to an individ, [17] then into interrogative *который "which"*, *чей "whose"* and narrative *картонный "cardboard (packaging)"*, *собачий "dog's"*.

G. Single-place and multi-place

Single-place taigens and yogens name one individ or one attribute of an individ *(шкаф "cabinet", коричневый "brown", открыть "to open once")*, while multi-place ones name a set of individs or a set of attributes of individs *(мебель "furniture", яркий "bright", открывать "to open twice")*. In complex cases, the power of Language Parts is determined procedurally: single-place taigens and yogens name a whole set, while multi-place ones name a whole set and some part of it, for example: *дерево "a tree"* is a single-place taigen (a part of a tree is not a tree, but a crown, trunk, or root), *лес "a forest"* is a multi-place (a part of a forest is also a forest). Compare with the morphological category of number: the lexeme *Афины "Athens"* has a plural form, while the lexeme *мебель "furniture"* has a singular form, but the former is a single-place taigen, while the latter is a multi-place one; in the lexeme *сады "gardens"*, the semantic and morphological categories coincide.

H. Intensities and extensiveness

Intensive taigens and yogens denote a whole set and some part of it, represented by a proper subset with the number of individs greater than, or equal to, one: *лес "forest"*, *мебель "furniture"*, *порошок "powder"*, *яркий "bright"*, *иметь "to have"* and **extensive** ones denote a whole set and some part of it, represented by a group of subsets with the number of individs greater than, or equal to, one: *trees, forests, powders, шагать "to step (many times)"*, i. e. intensive signs denote a set as a whole one, and extensive ones denote a set as a set. In complex cases, the procedure of minimization is used: if the power of the set denoted by the sign either cannot be reduced at all, i. e. it is impossible to single out a separate individ from the set, for example, *вода "water"*, or it is reduced to one, when the name of a set with the number of individs greater than one coincides with the name of one of its individs, for example, *пальто "coat"*, then the sign is intensive; if the limit of reducing the power of the set is two or more individs, then the sign is extensive, for example, *деревья "trees"*. For a number of variable yogens, intensity is equivalent to **iterativity**.

I. Positive and negative degree

An important distinguishing feature of yogens is their degree. The **first degree** yogens denote the individ's

[13] Compare with personal pronouns in the Greco-Latin classification of Speech Parts.

[14] **Numerical taigens** are a type of informational taigens that name numbers: *два "two"*, *десять "ten"*, *сто "one hundred"*, *тысяча "one thousand"*.

[15] Compare with cardinal and ordinal numerals, as well as with measured words in the Greco-Latin classification.

[16] The terms are borrowed from the paradigm of sentences, which as speech unites are divided by form into narrative and interrogative, by purpose into affirmative (containing an assertion) and imperative, by evaluation into positive and negative, by expressiveness into epic and exclamatory [48, p. 228].

[17] Compare with demonstrative, interrogative and possessive pronouns, as well as qualitative, relative and possessive adjectives in the Greco-Latin classification. Note that the traditional opposition of qualitative and relative adjectives does not stand up to criticism (see paragraph 1 of this article).

attribute *жёлтый "yellow"*, *сжимать "to squeeze"*, the **second degree** yogens denote the attribute of the attribute *тёмный "dark"*, *темно-коричневый "dark brown"*, *быстро "fast"*, *быстро бежать "to run fast"*, the **third degree** yogens denote the attribute of the attribute of the attribute *(ультра "ultra"*, *ультра/темно-коричневый "ultra dark brown"*, *очень "very"*, *очень быстро бежать "to run very fast")*. The second and third degrees of yogens can be positive or negative. The **second** and **third positive degree** yogens specify the meanings of the first degree yogens *ультра/темно-коричневый "ultra dark brown"*, *очень быстро бежать "to run very fast"*, and the first degree yogens specify the meanings of the **second** and **third negative degree** yogens *очень рад видеть "very glad to see"*, *очень любить бегать "to love running very much"*. The third positive or negative degree of yogens is determined syntagmatically: if the third degree yogen refers to the second positive degree yogen *очень быстро бежать "to run very fast"*, then it is positive, if the third degree yogen refers to the second negative yogen *очень любить читать "to love reading very much"*, then it is negative.

J. Resultativeness and non-resultativeness

Constant yogens are always resultative. A variable yogen is **resultative** if it denotes the subject's impact on object₁ or on object₂ with object₁ as a constant attribute of object₂ (deep impact); a variable yogen is **non-resultative** if it denotes the subject's influence on the closest environs of object₁ (i. e. on object₂), and object₁ is not a constant attribute of object₂ (superficial impact [18]). Such is, for example, the difference between the variable yogens *видеть "to catch sight of"* and *смотреть "to look"*, i. e. one can *look*, but not obtain a result *(to catch sight of)*. Graphically, the difference between a resultative and non-resultative yogens is depicted as follows (Fig. 1):

- 1 -– object;
- 2 – closest environs of object;
- 3 – subject;

(dotted line) denotes an object as a constant attribute of its closest environs.
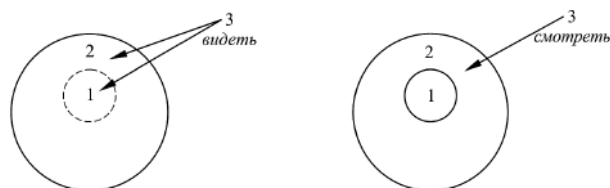

Figure 1.

[18] Deep impact changes the object, superficial impact can only move it.

K. Completeness and incompleteness

Variable yogens that name a completed or uncompleted process are classified as completed *брал "took"* and, accordingly, uncompleted *берёт "take"*.

L. Brevity and duration

Variable yogens denoting a short process *открыть "open"* are short, those denoting a long process *открывать "be opening"* are long.

The listed semantic categories of language allow us to formulate one-to-one rules in syntax and solve many important syntactic problems. For example, in relation to analytical applicative languages, such as Chinese, to derive the law of the Chinese syntax main line, which consists in alternating taigens with yogens in the role of the principal parts of a sentence in the basic sentence position, and to construct the Virtual String of Chinese syntax in the form of the complicated Chinese sentence generalized model, in which all its parts are realized [19].

III. Paradigm of Language Parts

A. The General Paradigm

The General Paradigm of taigens and yogens is graphically depicted in Fig. 2–3. Its purpose is to show the general in the hierarchy of Language Parts. For the sake of compactness and transparency of the presentation, the paradigmatic detailing of taigens is given using the example of constricted Parts of Language, and the paradigmatic detailing of yogens is given using the example of contracted Parts of Language; composite, complex, and contracted taigens are subdivided similarly to constricted ones, and composite, complex, and constricted yogens – similarly to contracted ones; the division of informational taigens and yogens is isomorphic to the division of physical ones.

B. The Special Paradigm

The Special Paradigm is focused on the difference in the hierarchical structure of taigens (Fig. 4) and yogens (Fig. 5). It is derived from the General Paradigm by introducing additional hierarchical levels that are characteristic only of taigens or only of yogens. **It was the Special Paradigm that made it possible to start creating a lexical analyzer based on semantic rather than morphological categorical features** within the framework of the scientific assignment and the research work of the same name "The Development of Artificial Intelligence Technologies Based on a Lexical Analyzer

[19] See: Hardzei A. Virtual String as a Syntactical Code of a Sentence (by the examples of the Chinese language). Proceedings of International Research Conference "Language, society and problems of intercultural communication". Grodno, Grodno State University Publ., 2007, pt. 2, pp. 349–358. (In Russian); Hardzei A. New Edition of the Virtual String of Chinese Syntax. In : The Paths of the Middle Kingdom. Minsk, Minsk State Linguistic Univ. Publ., 2025, iss. 12, pt. 1, pp. 25–37. (In Russian).
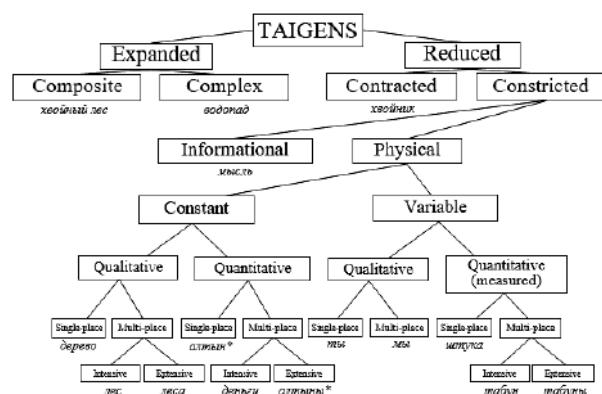
Figure 2. The General Paradigm of Taigens
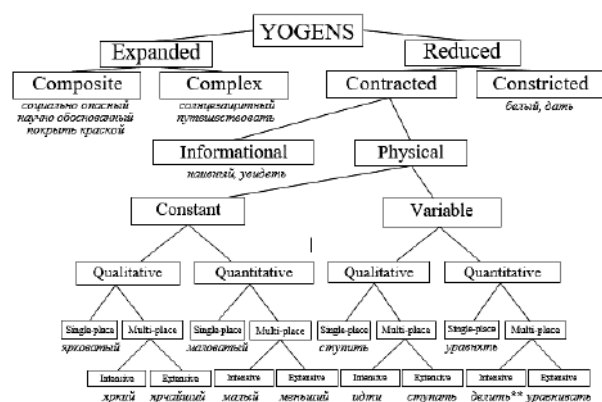*All original lexical borrowings are constricted

Figure 4. The Special Paradigm of Taigens
*Only when designating information individs.

Figure 3. The General Paradigm of Yogens
**Only when designating a physical process (делить пирог). Constricted physical quantitative single-place and multi-place extensive yogens in Russian have not been found; a contracted physical quantitative single-place variable is, for example, yogen уравнять, contrated physical variable quantitative multi-place extensive is yogen уравнивать.

Figure 5. The Special Paradigm of Yogens
**Constant physical modal yogens and single-place qualitative evaluative variable yogens have not yet been detected in the Russian language; examples of contant informational modal multi-place yogens are: вероятный, вероятнее.
***Only when designating the attributes of physical individs.

and Knowledge Bases" of the subprogram "Digital Space Technologies, Human, Social and State Security 1.1.7" (Resolution of the Presidium of the National Academy of Sciences of Belarus dated September 11, 2023, No. 28).

In the Special Paradigm, as in the General Paradigm, for the sake of compactness and transparency of presentation, the paradigmatic detailing of taigens is given using the example of constricted Parts of Language, and the paradigmatic detailing of yogens is given using the example of contracted Parts of Language; composite, complex and contracted taigens are subdivided similarly to constricted ones, and composite, complex and constricted yogens –similarly to contracted ones; the division of physical and informational taigens and yogens is isomorphic; proper taigens and yogens do not have stratification; constant physical modal yogens, and single-place qualitative evaluative variable yogens have not yet been detected in the Russian language, examples of constant informational modal yogens are: *возможный "possible"*, *возможнее "more possible"*, *вероятный*
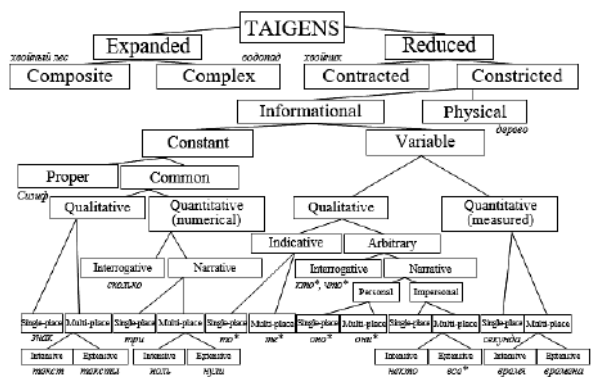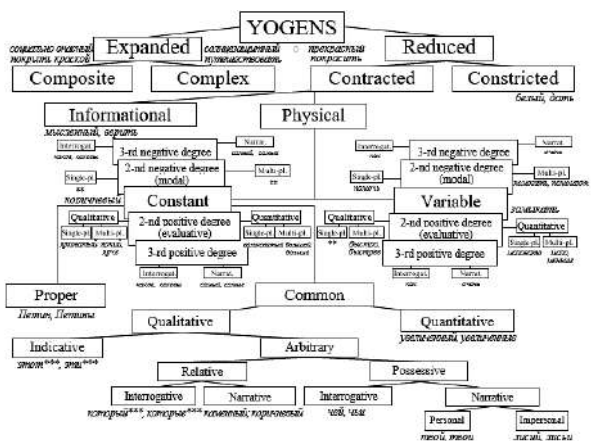
"probable", *вероятнее "more probable"*, etc.

Conclusion

The Special Paradigm of Language Parts is, in general, sufficient for the analysis of the syntax of applicative languages, such as Chinese. Its prospects for English and Russian are also beyond doubt, because by the stepwise division of the set of taigens and the set of yogens into subsets and subsets of subsets, any level of discreteness in the representation of linguistic units is achieved. In support of this, we will give several examples for further recursive detailing of the Special Paradigm as applied to the Russian language (Figs. 6–8). The time paradigm for variable yogens (Fig. 8) displays the **temporal meta-transformation** in the ontological aspect (*past* vs. *future*) [20]: *Сестра прочла книгу "The sister read the book" Сестра прочтёт книгу "The sister*

[20] They differ in the *point of singularity* on the time scale: before it is the past, after it is the future.

will read the book" and in the gnoseological aspect *(**present** vs. **future in the past**)* [21]: *Сестра читает книгу "The sister is reading the book" Сестра читала бы книгу "The sister would read the book"* [49, p. 29].
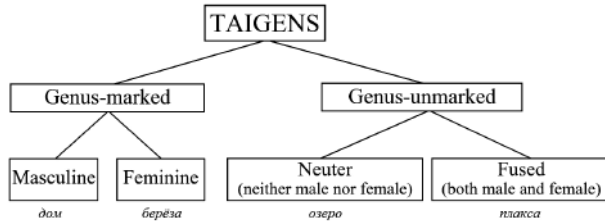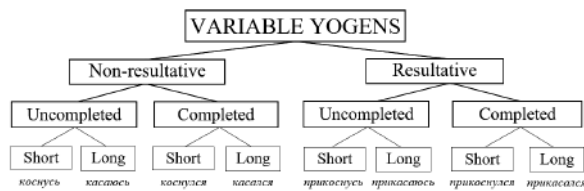

Figure 6. The Genus Paradigm of Taigens

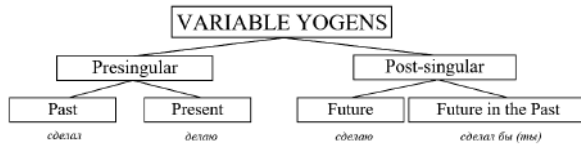Figure 7. The Aspect Paradigm of Variable Yogens

Figure 8. The Time Paradigm of Variable Yogens

In conclusion of the article, we note that in TAPAZ-2 the top-level ontology is specified constructively using an oriented Knowledge Graph with a total number of vertices $S_n \approx 8,2 \times 10^{245}$, built on a Semantic Classifier of 112 macroprocesses and a sheet of 32 roles of individs calculated by TAPAZ-2 algebra [50], which provides *a unique opportunity* to create a system for Natural Language Processing taking into account the already known properties of the semantic analyzer, which is based on an optimized version of the TAPAZ-2 Knowledge Graph (a graph suitable for processing by modern technical means [51], i. e. now is possible to solve the problem from top to bottom: from the semantic analyzer to the syntactic and lexical one. The advantage and effectiveness of this approach compared to existing methods are obvious [52]. **The implementation of this task will be a decisive step in the creation of an inventive machine**.

---

[21] The present is a singularity point on the time scale that has no extension. The interval is created by the observer, uniting the closest past with the closest future to facilitate orientation in time and space. The transfer of the future to the past is also dictated by subjective reasons, for example, the desire of the observer to emphasize the determinacy of events or to preserve the time plan of description.

References

[1] Tarski A. Logic, Semantics, Metamathematics; Papers from 1923 to 1938. Oxford, Clarendon Press; London, Oxford Univ. Press, 1956, xiv, 471 pp.

[2] Benveniste E. Problems in General Linguistics. Miami Linguistics Series. Florida, Coral Gables, University of Miami Press, 1971, no. 8, viii, pp. 1–317.

[3] Shikhanovich Yu. A. Introduction to Modern Mathematics. Moscow, Science, 1965, 376 pp. (In Russian).

[4] Mendelson E. Introduction to Mathematical Logic. 4th edn. London, Chapman & Hall Publ., 1997, x, 440 pp.

[5] Frege G. Begriffsschrift, a Formula Language, Modeled Upon That of Arithmetic, for Pure Thought. In : Heijenoort, J. van (ed.) From Frege to Gödel : A Source Book in Mathematical Logic, 1879–1931. MA, Harvard, Harvard University Press, 1967, pp. 1–82.

[6] Sikorski R. Boolean Algebras. Berlin, Heidelberg, New York, Springer, 1964, x, 237 pp.

[7] Kandrashina E. U., Litvintseva L. V., Pospelov D. A.: Knowledge Representation about Time and Space in Intelligent Systems. Moscow, Science, 1989, 328 pp. (In Russian).

[8] Cantor, G.: Foundations of a General Theory of Manifolds : A Mathematico-Philosophical Investigation into the Theory of the Infinite. In : Ewald, W. B. (ed.) From Kant to Hilbert : A Source Book in the Foundations of Mathematics. Oxford, Oxford University Press, 1996, vol. 2, pp. 878–920.

[9] Martynov V. V. Foundations of Semantic Coding. Experience of Knowledge Representation and Transformation. Minsk, European Humanitarian University Publ., 2001, 140 pp. (In Russian).

[10] Philosophy : Encyclopedic Dictionary. Ed. by A. A. Ivin. Moscow, Gardariki, 2004, 1072 pp. (In Russian).

[11] Rasiowa H., Sikorski R. The Mathematics of Metamathematics. 2nd edn. Waszawa, Polish Scientific Publishers, 1963, 501 pp.

[12] Lewis C. I. Implication and the Algebra of Logic. Mind, vol. 21, no. 84, 1912, pp. 522–31.

[13] Hilbert D., Ackermann W. Principles of Mathematical Logic ; translated from the German by Lewis M. Hammond, George G. Leckie, F. Steinhardt ; edited and with notes by Robert E. Luce. New York, Chelsea Publ. Co., 1950, ix, 172 pp.

[14] Anderson A. R., Belnap N. D. Entailment : The Logic of Relevance and Necessity, vol. I. Princeton, New York, Princeton University Press, 1975, 542 pp.

[15] Karaulov Yu. N., Molchanov V. I., Afanasyev V. A., Mikhalev N. V. Russian Semantic Dictionary : An Experience of Automatic Construction of Thesaurus : From Concept to Word. Moscow, Science, 1982, 566 pp. (In Russian).

[16] Ozhegov S. I. Dictionary of the Russian language. Moscow, Russian language, 1984, 816 pp. (In Russian).

[17] Galileo G. Dialogues Concerning Two New Sciences. Translated from the Italian and Latin into English by Henry Crew and Alfonso de Salvio. With an Introduction by Antonio Favaro. New York, Macmillan Co., 1914, 300 pp.

[18] Hausdorff F. Bemerkung über den Inhalt von Punktmengen. Mathematische Annalen, vol 75, 1914, pp. 428–434.

[19] Banach S., Tarski A. Sur la décomposition des ensembles de points en parties respectivement congruentes. Fundamenta Mathematicae, № 6, 1924, pp. 244–277.

[20] Lawvere F. W. Quantifiers and sheaves. Actes du Congrès International des Mathématiciens, vol. 1, Paris : Gauthier-Villars, 1971, pp. 329–334.

[21] Tierney M. On the Spectrum of a Ringed Topos. Algebra, Topology, and Category Theory : A Collection of Papers in Honor of Samuel Eilenberg, New York, San Francisco, London : Academic Press, 1976, pp. 189–210.

[22] Hilbert D. The Foundations of Geometry. Authorized translation by E. J. Townsend. Illinois, La Salle, The Open Court Publishing Co., 1950, 87 pp.

[23] Grammar of the Modern Russian Literary Language. USSR Academy of Sciences, Russian Language Institute ; resp. ed. N. Yu. Shvedova. Moscow, Science, 1970, 767 pp. (In Russian).

[24] Reichenbach H. The Philosophy of Space and Time. New York, Dover Publ., 1958, xvi, 295 pp.

[25] Shansky N. M, Tikhonov A. N. Modern Russian Language. In 3 parts. Part 2. Word Formation. Morphology. Moscow, Education, 1981, 270 pp. (In Russian).

[26] Kurilovich E. Lexical Derivation and Syntactic Derivation. In : Essays on Linguistics. Moscow, Foreign Literature, 1962, 456 pp. (In Russian).

[27] Germanovich A. I. Interjections of the Russian language. Kyiv, Radyanska shkola, 1966, 172 pp. (In Russian).

[28] Saussure F. de. Course in General Linguistics. Edited by Ch. Bally and A. Sechehaye in collaboration with A. Reidlinger. Translated from the French by W. Baskin. New York, Philosophical Library 1959, xvi, 240 pp.

[29] Jespersen O. The Philosophy of Grammar. London, George Allen & Unwin Ltd, 1951, 360 pp.

[30] Tesnière L. Éléments De Syntaxe Structurale. Paris, Librairie C. Klincksieck, 1959; English translation : L. Tesnière. Elements of Structural Syntax. Amsterdam / Philadelphia, John Benjamins Publishing Co., 2015, 698 pp.

[31] Fortunatov F. F. Selected Works. Moscow, Uchpedgiz, 1956, vol. 1, 450 pp. (In Russian.)

[32] 'It's a shame: Berners-Lee, The World-Web creator, criticizes Web3 and calls the idea nonsense [Internet edition]. Available at: https://devby.io/news/eto-pozor-sozdatel-vsemirnoi-pautiny-berners-li-raskritikoval-web3-i-nazval-ideu-chushu (accessed 18.03.2025). (In Russian).

[33] Winograd T. Understanding Natural Language. Cognitive Psychology, Academic Press, 1972, iss. 3, pp. 1–191.

[34] Berners-Lee T., Hendler J., Lassila O. The Semantic Web : A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. Scientific American, 2001, vol. 284, iss. 5, pp. 34–43.

[35] Hardzei A. Metasemantics of Language Categories. Second Readings Dedicated to the Memory of Professor V. A. Karpov. Minsk, Belarusian State University Publ., 2008, pp. 19–24. (In Russian).

[36] Bogdanov A. Essays in Tektology. English translation by George Gorelik. California, Seaside, Intersystems Publ., 1980, xvii, 265 pp.

[37] Kotarbinski T. Praxiology : An Introduction to the Sciences of Efficient Action. London, Oxford, Pergamon Press, 1965, 219 pp.

[38] Bertalanffy von L. General System Theory : Foundations, Development, Applications. New York, George Braziller, 1968, 289 pp.

[39] Boulding K. General Systems Theory – The Skeleton of Science. Management Science, 1956, vol. 2, no 3, pp. 197–208.

[40] Wiener N. Cybernetics or Control and Communication in the Animal and the Machine. Massachusetts, Cambridge, the M.I.T. Press, 1985, xvi, 212 pp.

[41] Ashby W. R. An introduction to cybernetics. New York, John Wiley & Sons Inc., 1956, ix, 295 pp.

[42] Neumann von J. The General and Logical Theory of Automata. In : Cerebral Mechanisms in Behavior : The Hixon Symposium. New York, John Wiley & Sons Inc., 1951, pp. 1–41.

[43] Hardzei A. Parts of Language and Parts of Sentence. In : Report Summary of International Scientific Conference on Semantic Potential of Language Units and its Implementation, October 20–21, 2021, Minsk State Linguistic University; ed. : L. M. Leshcheva [et al.]. Minsk, MSLU, 2021, pp. 9-11. (In Russian).

[44] Martynov V. V. Language Categories. The Semiological Aspect. Moscow, Science Publ., 1982, 192 pp. (In Russian).

[45] Hardzei A. Parts of Language and the Procedures of its Delineation. In : The Paths of the Middle Kingdom. Minsk, Belarusian State Univ. Publ., 2006, iss. 1, pt. 1, pp. 69–75. (In Russian).

[46] Isachenko A. V. On the Question of Structural Typology of the Vocabulary of Slavic Literary Languages. In : Slavia, Praha, 1958, roč. xxvii, seš. 3, s. 334–352. (In Russian).

[47] Kolmogorov A. N., Fomin S. V. Elements of the Theory of Functions and Functional Analysis. New York, Rochester Graylock Press., 1957, vol. 1, ix, 129 pp.

[48] Hardzei A. Linguistic Propaedeutics. In : Proceedings of the IV Republican Scientific Conference "Belarus in the Modern World". Minsk, Republican Institute of Higher Education, 2005, pp. 226–229. (In Russian).

[49] Hardzei A. New Edition of the Virtual String of Chinese Syntax. In : The Paths of the Middle Kingdom. Minsk, Minsk State Linguistic Univ. Publ., 2025, iss. 12, pt. 1, pp. 25–37. (In Russian).

[50] Hardzei A. Theory for Automatic Generation of Knowledge Architecture : TAPAZ–2; transl. from Rus. I. M. Boyko. Rev. English edn. Minsk, Republican Institute of Higher Education, 2017, 50 pp.

[51] Hardzei A., Udovichenko A. Graph of TAPAZ–2 Semantic Classifier. In : Open Semantic Technologies for Intelligent Systems. Minsk, Belarusian State Univ. of Informatics and Radioelectronics, 2019, iss. 3, pp. 281–284.

[52] Natural Language Processing [Internet edition]. Available at: https://habr.com/ru/companies/vk/articles/358736/ (accessed: 18.03.2025). (In Russian).

ЛИНГВИСТИЧЕСКОЕ ОБОСНОВАНИЕ СОЗДАНИЯ ЛЕКСИЧЕСКОГО АНАЛИЗАТОРА НА СЕМАНТИЧЕСКОЙ ОСНОВЕ

Гордей А. Н.

Подробно рассматриваются логические, семантические и метасемантические парадоксы, возникающие при смешении теории и метатеории во время описания иерархических систем, объясняется их значение для лингвистики, приводятся примеры внутренней противоречивости учения о частях речи, вызывающей масштабные ошибки при обработке данных на естественном языке, обосновывается создание лексического анализатора в парадигме частей языка и теории автоматического порождения архитектуры знаний (ТАПАЗ–2).