Semantic analysis of Natural Language.

overview by Poroshin V.A.

INTRODUCTION

I think, everyone understands role of Natural Language (NL) as a tool to represent information. For our computer age it is quite obvious and extremely important to retrieve information from NL or make it *processable* by computer. I'm using word *processable* instead of more popular and clever one – *understandable*. But here is a big difference. Computer is not able to understand not only NL but any other data including such an artificial language like PASCAL or Java. It might look like nonsense but from my point of view *understanding* is a mental process, i.e. how human beings recognize objects (mental and physical) and links between them. Since computer does not have a human mentality, so it cannot understand by definition.

Of course, NL processing (NLP) is a general problem and to be more specific we can separate it by categories according to increasing level or complexity of such processing:

- Morphology and morphological processing
- Syntax and syntactical processing
- Semantics and semantic processing

Morphology is a subdiscipline of linguistics that studies word structure. During morphological processing we are basically considering words in a text separately and trying to identify *morphological classes* these words belong to. One of the widespread task here is *lemmatizing* or *stemming* which is used in many web search engines. In this case all morphological variations of a given word (known as word-forms) are collapsed to one *lemma* or *stem*.

Syntax as part of *grammar* is a description of how words grouped and connected to each other in a sentence. There is a good definition of syntax for programming languages: "… syntax usually entails the transformation of a linear sequence of tokens (a token is akin to an individual word or punctuation mark in a natural language) into a hierarchical syntax tree". Later we will see that the same definition also can be used for NL. Main problems on this level are: *part of speech tagging* (POS tagging), *chunking* or detecting *syntactic categories* (verb, noun phrases) and *sentence assembling* (constructing syntax tree).

Semantics and its understanding as a study of meaning covers most complex tasks like: finding synonyms, *word sense disambiguation*, constructing *question-answering systems*, translating from one NL to another, populating base of knowledge. Basically one needs to complete morphological and syntactical analysis before trying to solve any semantic problem.

Formalization of NL leads us to solutions of all these problems. There are a lot of theories and opinions about how to do it. But the only proof of efficient solution here can be a working computer program. Such a program exists. It can analyze Russian newspapers with a very high precision[*]. And this program is based on Prof. Tuzov V.A. theory shortly described here[†].

[*] No evaluations so far. I'll talk about reasons on seminar.

[†] Prof. Tuzov was my supervisor during several years of my study in Saint-Petersburg State University on faculty Applied Mathematics and Control Processes. This overview is based on his books and my own researches in this area.

CHAPTER I. BASIC FORMALIZING PRICNCIPLES OF NATURAL LANGUAGE.

§1. Basic structure of Natural Language.

Problem of formalizing human language as a part of Computational Linguistics originated from 1950 then American scientists tried to translate different NL to English. Since that time, there were a lot of discussions about general structure of NL and ways how it can be processed by computer. It is quite obvious that in order to solve complex NLP tasks, especially related to semantic analysis, we need formal representation of language i.e. *semantic language.* The basis of such semantic language is sequence of simple and mathematically accurate principles which define strategy of its construction:

Thesis 1. *Language is algebraic system {f1, f2, ... , fn, M), where* f_i *is a* ***basic function*** *and M is a structure of a given language (****basic concepts****).*

This thesis gives us principal answer how language is organized and it claims that there is one universal grammar for all languages: natural and artificial.

Thesis 2. *Adequate grammar describers each sentence as a structural superposition of functions.*

If some grammar describes sentences using another way then it can not be adequate for NL.

Thesis 3. *Grammar of specific language is a concrete definition of universal grammar.*

If two languages are semanticly equivalent then the only difference is in names of functions and arguments.

Thesis 4. *Each part of speech has well defined role to organize syntactical structure of sentence.*

Nouns as arguments of functions define data structure of language *M.* Adjective is simple function takes nouns as a parameter. Verb is function mainly on nouns. Adverb is function on verbs and so on.

Thesis 5. *Grammar is linked to semantics of language and it is represented as semantic dictionary.*

Each word must be defined as some aggregate of semantic formulas. The source of these definitions is a semantic dictionary. The abstraction degree of formula depends on many reasons but the less abstract formula is, so the more exactly computer can react to words linked with it.

Corollary 1. *It is better to define word in a very abstract way instead of skip it.*

Corollary 2. *Syntactic structure of sentence is adequate reflection of its semantic structure.*

Thesis 6. *Learning language is a process of building and replenishment of semantic dictionary.*

Then human being learns language he/she studies not only words but mainly functions linked with them. You can use some word only if you know how to use other words with it.

Corollary 1. *Semantics is a basis of syntax learning.*

Corollary 2. *Necessary condition for any communication is a similarity of semantic dictionaries.*

Thesis 7. *Language does not make a separation between physical or mental concepts of the world.*

Thesis 8. *Language is an idea that we can not pronouncedly define but there is no language that we can not formalize.*

In other words these claims are nearly equal to following two assertions:

1. Word of any natural language is a name of function *f(x1, ..., xn)* which is linked with it and called its semantics. Word gets every concrete meaning only if we consider its function with *concrete* values of *x1, ..., xn*. This is the same idea as in mathematics. For instance, function *y = sin(x)* gets concrete values of *y* only after calculating with some concrete value of *x*.
2. Sentence is a united superposition of functions, i.e. sentence is a expression in mathematical point of view. For example, *sin(x+y)* is a sentence but *sin(x* is not. And the same is: "*He took his book*" is a sentence and "*He took his*" is not. The meaning of sentence is calculated as a process of execution of superposition (§4).

So, each word of sentence is a superposition *F* of *basic functions* $f_1,\ldots,f_n$, where arguments of *F* are *basic concepts* of language.

Let us consider example where meaning of such theoretical constrictions can be clarified. We need to formalize the following sentence:

Mother bought a new dress.

Skipping a lot of details on this point, from *semantic dictionary* we'll get corresponded entries for words:

Word	***Lemma***	***Formula***
Mother	*Mother*	*12413511(genitive)*
bought	*buy*	*PerfOper01(nominative, BUY(accusative, from accusative, for accusative))*
new	*new*	*203021(nominative/accusative)*
dress	*dress*	*11363(genitive)*

Example 1.1

Number of types *12413511, 203021, 11363* are numbers in hierarchy of basic concepts (§2) and *PerfOper01* illustrates superposition of two basic functions (§3) *Perf* and *Oper01*. Arguments of functions show what kind of grammatical types can be connected. For instance, word *mother 12413511(genitive)* belongs to class *12413511* and can connect to itself some genitive (possessive) noun or pronoun. Verb *buy* can connect four arguments with different grammatical types.

After we have complete description of all words in a sentence there are several ways to assemble them. The final results we want is a complete superposition of function formalizing this sentence. In our case:

PerfOper01(Mother$12413511 (#), BUY(new$203021(dress$11363(#)), #, #)) or

PerfOper01 (mother(), BUY(new(dress))))

Also we can read this formula in a literal sense: *Mother did a buy of new dress*. It is so because of basic function *Oper01(x,y)* that means *x is doing y* and function *Perf* that leads us to past tense of this action.

Although, construction such semantic formula form ex. 1.1 is a simple task for human being it is not trivial for computer. There are different approaches but one of them needs to find *central word* in a sentence. Under *central word* we understand word (basically it is unique in a sentence and it is a verb) on which all syntactical structure of the sentence is based on. Clearly, in our case this word is verb *bought*.

After we have found a central word we can use its semantic formula to join corresponded arguments: first argument is a nominative or subject *Mother*. Word *new* is joined with *dress* and this constriction *new(dress)* with grammatical type of noun *dress* is acting as second argument of verb's formula. All words are linked and sentence analysis is complete.

This trivial example hides all details and problems we can face on in a real NL text analysis. But basic idea of what can be done and how will remain the same.

Let me point out that this idea of semantic analysis is language independent. Russian language (for which actually this theory was originally developed) has more inflectional nature then English. Six cases for nouns instead of three (or even two if we combine object and subject as one for English nouns) cases in English make formulas a little bit different:

Word	***Lemma***	***English match***	***Formula***
Мама	*Мама*	*Mother*	*12413511(genitive \ dative)*
купила	*купить*	*bought*	*PerfOper01(nominative, ПОКУПКА(accusative \ genitive, from genitive\ where, dative \ for))*
новое	*новый*	*new*	*203021(SOMETHING$1)*
платье	*платье*	*dress*	*11363(genitive)*

Example 1.2

More cases in Russian language produce more possibilities to describe arguments of formulas in grammatical point of view. Instead of only two distinguishable English cases for nouns Russian language with six cases can be more exactly described in formulas and probably this makes it easier for semantic analysis. It might disagree with common opinion that Russian language is more complex then English. But more concrete description for words produces more precise analysis since most of the alternative will be dropped as irrelevant (§4).

Chinese language, mostly distant and strange from European point of view, is not so far from English after more close understanding of its structure.

Word	***English match***	***Formula***
母亲	*Mother*	*12413511(genitive)*
买了	*bought*	*PerfOper01(nominative, BUY(accusative, from accusative, for accusative))*
新	*new*	*203021(nominative/accusative)*
衣服	*dress*	*11363(genitive)*

Example 1.3

Only two cases and strict position of words in a sentence make it more close to English then to Russian language.

At last Finnish language also can be formalized in our way (home work ;) :

Example 1.4

§2. Hierarchy of basic concepts of NL

Let's consider some problematic examples:

Dense bought a new mother. *(2.1)*
Mother bought a new dress from silk. *(2.2)*
Mother bought a new dress from Chinese store. *(2.3)*

Sentence (2.1) is syntactically correct and we still can construct corresponded formula for it:

PerfOper01 (new(dress), BUY(mother()))

Of course, we can assume that person who wrote it meant what he/she meant but it is a good idea to recognize this kind of nonsense and at least produce a warning that sentence is potentially not correct in semantic point of view.

More deep and serious problems are in examples (2.2) and (2.3). First of all, the same construction of preposition *from* with noun has different meaning and grammatical type. *Dress from silk* means *dress made from silk* but *dress form Chinese store* is a dress taken from concrete place. Syntactically and based only on morphological information we can not separate these two cases. What we need is semantic specification for arguments in our formulas from example 1.1. This additional information is a hierarchy of basic concepts.

Basic concept is a word of natural language which meaning can not be expressed using other more simple concepts. This means that vocabulary of NL can be partitioned into basic and non-basic clusters. Non-basic words are expressed using basic concepts and basic functions and called **derived words**. Let me illustrate this:

Word	***Part of speech***	***Formula***
fire	*noun*	$1225(Z1: *genitive*)
burning	*noun*	Lab_o(Z1: genitive, Z2: FIRE$1225 accusative)
flare	*noun*	Lab_o(Z1: genitive, Sing(Z2: FIRE$1225 accusative))
to burn	*verb*	Caus(Z1: reason, Lab(Z2: nominative, Z3: FIRE$1225 accusative))

Here first word *fire* in a sense of *hot glowing gases during burning of something* is a basic concept. There is no any description for it except position in the hierarchy (class number $1225) and arguments Z_i it can join. Other words in our example are derived form it. I.e. we can read their formulas as formal explanatory dictionary:

burning – something Z1 is exposed by something from class FIRE$1225;
flare – something Z1 is exposed by element of something from class FIRE$1225;
to burn – something Z1 is a reason that something Z2 is exposed by word from class FIRE$1225;

Basic concepts are organized in hierarchical tree. This hierarchy is based on rules:

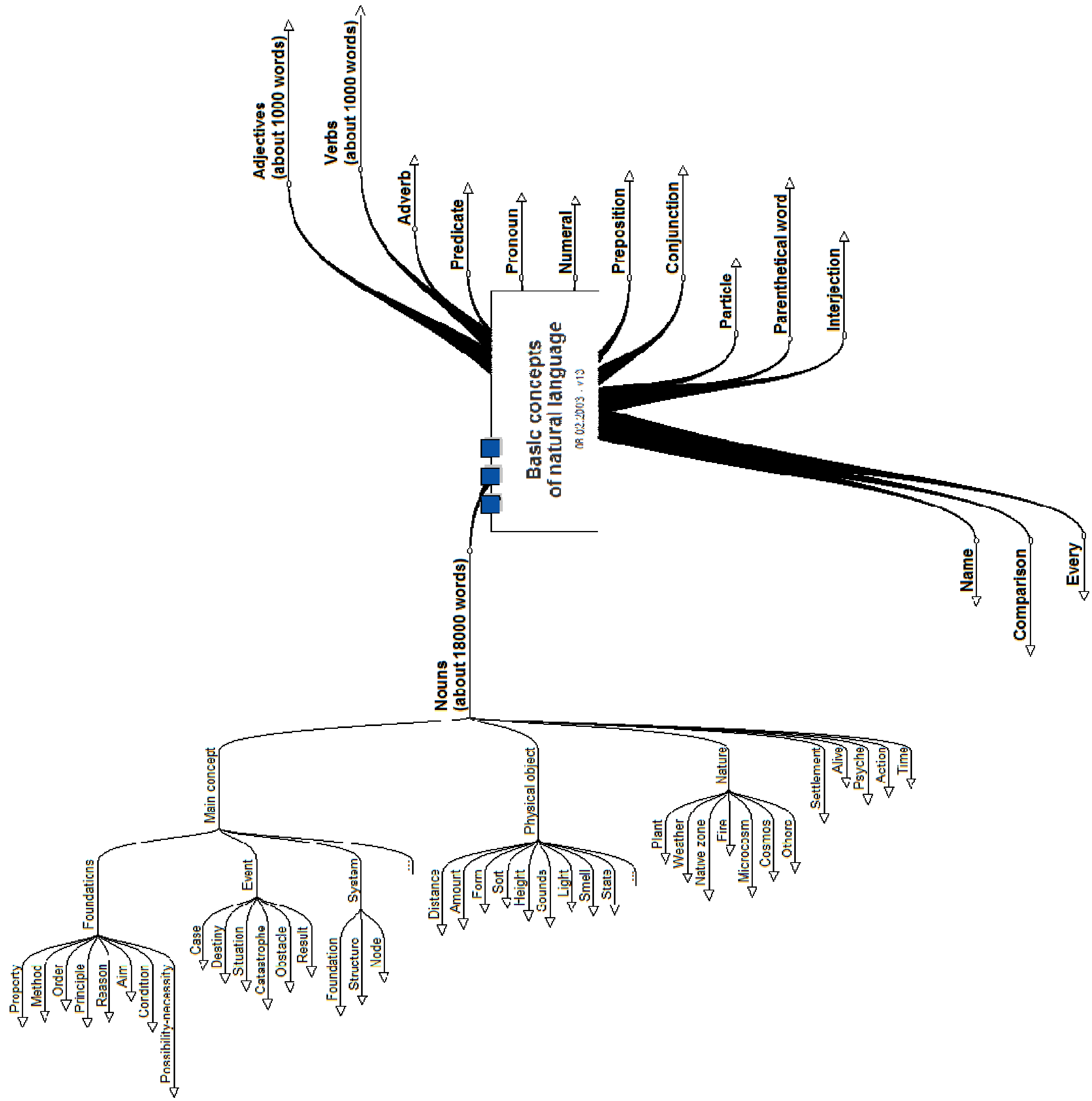
1. All elements of class must have the same semantic properties which are inherited from superclass. Elements of class do not only inherit all semantic properties of superclass but also they can have its own characteristic. The Root class of all branches is called "SOMETHING". Any class can have one or some subclasses but too much fragmentation is not expedient.
2. If one element of any class can be used as an argument in some semantic formula, then all the rest elements of this class can be arguments in this formula. The same is true for subclasses of given class, but the opposite assertion is not correct.
3. Each class has its own identification number that includes all superclasses of given class.

Hierarchy sample:

Class number	***Class name***
$0	Phraseology
$1	Noun
$110	Noun AO (Abstract Object) Idea
$1100/01	Noun AO Idea => Abstract-Concrete
$1100/02	Noun AO Idea => Defined-Non-defined
$1101	Noun AO Idea Property
$11011	Noun AO Idea Property Quality
$1101111	Noun AO Idea Property Quality Concrete
$110112	Noun AO Idea Property Quality Category
$110113	Noun AO Idea Property Quality Class
$11014	Noun AO Idea Property Role
$11016	Noun AO Idea Property Estimation
$1102	Noun AO Idea Method
$11021	Noun AO Idea Method Instrument
$1103	Noun AO Idea Order
$11031	Noun AO Idea Order Harmony
$11032	Noun AO Idea Order Ordinance
$11033	Noun AO Idea Order Regulations
$1104	Noun AO Idea Principe
$1105	Noun AO Idea Reason
$12	Noun PO (Physical Object)
$12/001	Noun PO => Distance
$12/0011	Noun PO => Distance Unit
$12/0012	Noun PO => Distance Measuring
$12/002	Noun PO => Direction
$12/00201	Noun PO => Direction Left-Right
$12/00202	Noun PO => Direction Nearby-Faraway
$12/00203	Noun PO => Direction Inside-Outside
$12/00204	Noun PO => Direction In_front_of-behind
$12/00205	Noun PO => Direction Higher-Lower
$12/00206	Noun PO => Direction Here-There
$12/004	Noun PO => Sort
$12/005	Noun PO => Growth
$12/0051	Noun PO => Growth Increasing
$12/006	Noun PO => Sound
$12/0060	Noun PO => Sound Silence
$12/0061	Noun PO => Sound Concrete
$12/0062	Noun PO => Sound of_Animals
$12/0063	Noun PO => Sound Voice
$12/00631	Noun PO => Sound Voice Concrete
$12/006401	Noun PO => Sound Quiet-Loud
$12/007	Noun PO => Light
$12/0071	Noun PO => Light Sparkling
$12/0072	Noun PO => Light Shadow
$120	Noun PO Quantity
$1200	Noun PO Quantity Number
$12001	Noun PO Quantity Number Digit
$122	Noun PO Nature
$122/1	Noun PO Nature Weather

$122/101	Noun PO Nature Weather Climate
$122/11	Noun PO Nature Weather Wind
$122/12	Noun PO Nature Weather Water
$122/13	Noun PO Nature Weather Light
$122/14	Noun PO Nature Weather Ice
$122/15	Noun PO Nature Weather Precipitation
$123	Noun PO Settlement
$12300	Noun PO Settlement Countries
$123001000	Noun PO Settlement Countries Name
$12301	Noun PO Settlement Province
$12302	Noun PO Settlement Town
$124	Noun PO Alive
$124/0	Noun PO Alive Life
$124/00	Noun PO Alive Life Perception
$124/001	Noun PO Alive Life Perception Vision
$124/002	Noun PO Alive Life Perception Hearing
$124/003	Noun PO Alive Life Perception Smelt
$124/004	Noun PO Alive Life Perception Touch
$124/2	Noun PO Alive Diseases
$124/20	Noun PO Alive Diseases Health
$124/201	Noun PO Alive Diseases Health Illness
$124/202	Noun PO Alive Diseases Health Convalescence
$124112	Noun PO Alive Human People
$12411200	Noun PO Alive Human People Organization
$12411201	Noun PO Alive Human People Nation
$12411202	Noun PO Alive Human People Military
$12411203	Noun PO Alive Human People Группа
$13	Noun Psyche
$130	Noun Psyche Soul
$1300	Noun Psyche Soul Senses
$130001	Noun Psyche Soul Senses Temperament
$141	Noun Art
$1411	Noun Art Literature
$142	Noun Science
$142/3	Noun Science Methods
$142/4	Noun Science Special_definitions
$15	Noun Action
$150	Noun Action Modification
$1501	Noun Action Modification Transformation
$16	Noun Time
$16/01	Noun Time => Temporary-Eternal
$16/02	Noun Time => Early-Late
$702	Pronoun
$703	Adverb
$711	Preposition
$712	Conjunction
$713	Particle

Here is a picture of top-level hierarchy of basic concepts:



Using this classification now we can clarify description of formulas and solve our problems in examples 2.1, 2.2 and 2.3:

Word	***Lemma***	***Formula***
Mother	*Mother*	*12413511 (genitive)*
bought	*buy*	*PerfOper01(**ALIVE**$124 nominative, BUY(accusative, **PLACE** from accusative, for accusative))*
new	*new*	*203021(nominative/accusative)*
dress	*dress*	*11363(genitive)*

Example 2.4

Parameters in formula of verb *bought* now must obey additional restrictions: subject who bought is from class ALIVE$124 and its subclasses. *Dress* is not ALIVE (its class number does not match class 124), so it can not buy anything but ALIVE *mother* (*12413511* match 124) can. Therefore sentence *"Mother bought a new dress"* is semantically correct while sentence "*Dress bought a new mother"* is not correct.

Word	***Lemma***	***Formula***
Mother	*Mother*	*12413511 (genitive)*
bought	*buy*	*PerfOper01(**ALIVE**$124 nominative, BUY(accusative, **PLACE** from accusative, for accusative))*
new	*new*	*203021(nominative/accusative)*
dress	*dress*	*11363(genitive)*
from	*from*	*1212(**MATERIAL**$1212 nom./acc.)//019*
silk	*silk*	*1212111*

Example 2.5

Preposition *from* has several meanings and descriptions in semantic dictionary but the only correct formula for sentence (2.2) is *1212(**MATERIAL**$1212 nom./acc.)*. *Silk* as a subclass of MATERIAL (number of class *silk 1212111* has number of superclass *material 1212*) is joined with our correct alternative of prep. *from.*

Word	***Lemma***	***Formula***
Mother	*Mother*	*12413511 (genitive)*
bought	*buy*	*PerfOper01(**ALIVE**$124 nominative, BUY(accusative, **PLACE**#123 from accusative, for accusative))*
new	*new*	*203021(nominative/accusative)*
dress	*dress*	*11363(genitive)*
from	*from*	*123(**PLACE**$123 acc.)*
Chinese	*Chinese*	*CHINA$1231000(**SOMETHING**$1 nom./acc.)*
store	*store*	*$123402(…)*

Example 2.6

Meaning of preposition *from* in this example is different from previous one. Expression *Chinese store* has a grammatical type of noun and formula *CHINA$1231000($123402(…)). Chinese store* belongs to class *PLACE* and we choose corresponded meaning of *from* as *123(**PLACE**$123 acc.).* Finally, we have:

PerfOper01(Mother$12413511 (#), BUY(new$203021(dress$11363(#)), from$123(CHINA$1231000(store$123402(…))), #))

§3. Derived words of NL

Most of words in NL are not basic concepts. Their meaning can be expressed by more simple units: basic concepts and basic functions. Dictionary of all words of particular NL described in this manner is a **semantic dictionary**. Methods of how such a dictionary can be build are different from one language to another. Inflectional nature of Russian language allows partially automate this process. For non-inflectional languages like English other ways must be found. But hierarchy of basic concepts (only with some minor modifications) and all basic functions we can use the same in any other NL.

Basic function is a description of relationships between its arguments. Class of basic functions is some kind of hierarchy. On the top of it there are more abstract and universal functions. Lower layers consist of more concrete ones. All set of basic functions was appeared as a result of hand-work analysis of Russian language during long period. They are universal in terms of their usage area and portability to other languages.

Here we define all basic functions. Using them it is possible to build different superposition and try to translate them into new concepts of NL or, inside out, take some word of language and try to find suitable formula for it.

Here is a list of all basic functions:

Function	**Description**	***Function***	**Description**
And(x,y)	x and y	*Loc(x,y)*	x situated in y
Anti(x)	antonym of x	*Magn(x)*	x higher of norm
Apart(x,y,z)	x separates y from z	*Mult(x)*	multiset of x
Aspect(x,y)	x looks like y	*Ne(x)*	negation of x
Bon(x,y)	x is good for y	*Norm(x)*	x is in norm
Caus(x,y)	x is reason of y	*Oper(x,y)*	x performs y
Cond(x,y)	x is in conditions y	*Or(x,y)*	x or y
Cont(x)	x is continuing	*Percep(x,y)*	x percepts y
Content(x,y)	x contents y	*Plus(x)*	x is increasing
Control(x,y)	x controls y	*Poss(x)*	x is possible
Copul(x,y)	x is y	*Prepar(x,y)*	x prepares to y
Cor(x)	x is true	*Rel(x,y)*	x relatively to y
Degrad(x)	x is degrading	*Repet(x)*	x is repeating
Depend(x,y)	x depends from y	*Result(f)*	result of f
Emo(x)	inclination of x	*Role (x,y)*	x is in y role
Fact (x)	x is fulfilled	*Sing(x)*	element of set x
Fin (x)	x is ended	*Stato(x)*	status of x
Func(x)	x occurs	*Temp(x,y)*	x is time of y
Hab(x,y)	x has y	*Usor(x,y)*	x is used for y
Humaro(x)	mood of x	*Var(x,y)*	x is transforming into y
Incep(x)	x is starting	*Perf(x)*	perfect form of x
Intemp(x)	while x		
Kon(x,y)	x together with y		
Kontr(x,y)	x opposites to		
Lab(x,y)	x exposes y		

It is very important that this set of rules is finite and powerful enough to reflect all the richness of NL. For example, nearly 40 per cents of verbs in Russian language are *course* verbs. Therefore the more wonderful and powerful function is *Caus(x,y)*.

More examples:

word	***formula***	***translation of a formula***
author	*PerfCaus_o1(HUMAN$1241, IncepFunc(Z1))*	*HUMAN* who is a reason that something *Z1* has starting to occur
agglomeration	*Caus_o(#,IncepFunc(Z1: AGGLOMERATE$12125))*	something is a reason that some *AGGLOMERATE* is starting to occur
activation	*Caus(Z1,IncepCopul(Z2,ACTIVE$1100/12))*	something *Z1* is a reason that something *Z2* is starting to be *ACTIVE*
amnesty	*PerfCaus_o(#,FinHab(Z1,GUILT$131352))*	something is a reason that *Z1* has ended to have a *GUILT*

Here *AGGLOMERATE$12125, ACTIVE$1100/12* and *GUILT$131352* are basic concepts.

§4. Algorithm of semantic analysis

Execution of functions on a level of grammatical (syntactical) types is also called **pseudo-execution**. It is used in programming languages to analyze concrete-dependent conditions. Concrete-independent analysis is building superposition of functions while pseudo-execution is checking correctness of analyzed program. If superposition is correctly executed then program itself syntactically correct; if any function during execution becoming undefined then program is not correct. Although making analogies between natural and artificial languages is quite dangerous, nevertheless semantic analyzer in principle is working in the same way as compiler.

But there are several quite important differences. First of all, there are plenty of details to fight with. Secondly, if arguments for some function are not right then it does not mean (in programming languages it does) that analyzed sentence is incorrect. It might be so that these arguments must be linked with other word in the sentence. In the third place, and maybe most important, the same word-form may have different part of speech alternatives. And it is especially so in English language. But in spite of serious differences in a basis of semantic analysis there is a pseudo-execution.

Semantic analyzer coordinates of linking of two words. One of these word is treated as function, other – as its argument. Values of arguments are class numbers with grammatical types, for instance, **$1 nominative**, **$134 where, $167 what_kind_of**. Needed information for analyses is taken form semantic dictionary beforehand.

Let's consider an example of how algorithm of semantic analysis works for sentence:

Bill bought a book from Peter. (4.1)

First of all we need to complete grammatical analysis:

alter. num	***alter.***	***grammatical analysis***	***sense description***
Bill			
001	BILL	noun, proper name; singular form, subject	*somebody's name*

002	BILL	noun, common noun; singular form, subject	*a piece of paper that shows how much you owe sb for goods or services*
003	BILL	noun, common noun; singular form, subject	*in parliament a written suggestion for a new law*
004	BILL	verb; present simple	*to send sb. a bill for sth.*
…[‡]	…	…	…
bought			
001	BUY	verb; past simple	*to obtain sth by paying money for it*
002	BUY	verb; past simple	*to believe that sth is true, especially sth that is not very likely*
…	…	…	…
book			
001	BOOK	noun, common name; singular form, object	*a set of printed pages that are fastened inside a cover so that you can turn them and read them*
002	BOOK	noun, common name; singular form, object	*the written records of the financial affairs of a business*
…	…	…	…
from			
001	FROM	prep.	*used to show who sent or gave sth/sb*
002	FROM	prep.	*used to show where sb/sth starts*
003	FROM	prep.	*used to show when sth starts*
004	FROM	prep.	*used to show the material that sth is made of*
…	…	…	…
Peter			
001	PETER	noun, proper name; singular form, object	*somebody's name*

On this level we can drop several alternatives which do not satisfy morphological structure of English sentence. So, word *Bill* is at the beginning of sentence and after it we have a verb *bought*. Clearly, there is no way to be word *Bill* a verb. It means we can remove all verb descriptions of *Bill* form our list of alternatives. Article ***a*** before word *book* shows that this *book* is a noun and any other descriptions must be avoided. All that kind of rules depend on particular language and must filter out as many alternatives of sentence's words as possible.

On the second stage, for all correct alternatives we retrieve entries from semantic dictionary:

alter. num	***alter.***	***Formula***	***sense description***
Bill			
001	BILL	1241301000()	*somebody's name*
002	BILL	$1214(Z1: what_kind_of, Z2: from\for\...)	*a piece of paper that shows how much you owe sb for goods or services*
003	BILL	$1430(Z1: what_kind_of, Z2: about)	*in parliament a written suggestion for a new law*
…	…	…	…
bought			
001	BUY	PerfOper01(Z1: ALIVE$124 nominative, BUY(Z2: accusative, Z3: PLACE$123 from accusative, Z4: for accusative))	*to obtain sth by paying money for it*
002	BUY	BELIEF$14201(Z1: in_what)	*to believe that sth is true, especially sth that is not very likely*
…	…	…	…
book			
001	BOOK	$1430(Z1: about_what, Z2: AUTHOR$1241326 by)	*a set of printed pages that are fastened inside a cover so that you can turn them and read them*
002	BOOK	RECORD$1400611(Z1:	*the written records of the financial affairs of a*

[‡] Other alternatives are omitted for simplicity

		about_what,Z2:COMPANY$123411 by)	*business*
…	…	…	…
from			
001	FROM	ALIVE$124(Z1: ALIVE$124)	*used to show who sent or gave sth/sb*
002	FROM	123(PLACE$123)	*used to show where sb/sth starts*
003	FROM	1613 (Z1: TIME$1613)	*used to show when sth starts*
004	FROM	1212(MATERIAL$1212)	*used to show the material that sth is made of*
…	…	…	…
Peter			
001	PETER	1241301000()	*somebody's name*

Now we need only to pick up correct alternatives and join them into one superposition. In §2 I mentioned about *central word* algorithm. If we have a good guess which word is central one in a sentence we can significantly simplify process of analysis. For our example make a selection of this word is easy – one verb in a sentence most probably is a central word. For more complex sentence we could have try to treat all its verbs one by one as central. Or even do a search though all combinations of alternatives. Of course, this can be computationally expensive and slow. But it works in any way since there is a correct combination of alternatives to complete sentence analysis while this sentence is correct and our semantic dictionary is well constructed. Also this brute force approach is the only method we can use if there is no additional information about structure of analyzed language.

When central word is founded we need only to link its arguments with other words in a sentence. Let's consider first alternative of verb bought (marked as *bougth//001*):

PerfOper01(Z1: ALIVE$124 nominative, BUY(Z2: accusative, Z3: PLACE$123 from accusative, Z4: for accusative)). (4.2)

Noun *Bill* as a subject is its first argument Z1. And the only first alternative of *Bill*//001 is good for us because *Bill//001* belongs to class ALIVE$124. Other *Bill*'s al alternatives are not from class ALIVE$124, so there is no need to consider them.

Parameter *Z2* in (4.2) does not have any class description (you can buy anything). It can join *book//001* as well as *book//002*. At this point we can fork process of analysis and try to complete it for both cases of word *book*. It produces two separate and correct results with meanings: 1) *Bill bought "a set of printed pages for reading" from Peter* and 2) *Bill bought "written records of financial affaires" from Peter*. We understand that the second meaning is quite exotic but still can be true. From practical point of view it is better to avoid such a unusual results of semantic analysis. For example, we can follow the rule that if there is no additional information from context let's pick up only most probable variants between correct alternatives. As a result we have only one alternative *book//001*.

Structure *from Peter* has only one interpretation: ALIVE$124(Z1: PETER$1241301000). It is obvious since there is only one alternative of *Peter* and it is from class ALIVE$124. And only preposition *from//001* can join this argument. Now *from Peter* has grammatical type *from accusative*. And we have such an argument Z3 in (4.2). All words in are linked and sentence (4.1) correctly formalized.

Any other alternatives of verb *bought* will not lead us to complete sentence analysis: we simply won't join all words in one single formula.