ITHEA

International Journal INFORMATION TECHNOLOGIES & KNOWLEDGE

2009 Volume 3 Number 1

International Journal
INFORMATION TECHNOLOGIES & KNOWLEDGE
Volume 3 / 2009, Number 1

IJ ITK is official publisher of the scientific papers of the members of the ITHEA International Scientific Society

IJ ITK rules for preparing the manuscripts are compulsory.
The **rules for the papers** for IJ ITK as well as the **subscription fees** are given on ***www.ithea.org***
The camera-ready copy of the paper should be received by e-mail: info@foibg.com.
Responsibility for papers published in IJ ITK belongs to authors.

General Sponsor of IJ ITK is the **Consortium FOI Bulgaria** (www.foibg.com).

International Journal "INFORMATION TECHNOLOGIES & KNOWLEDGE" Vol.3, Number 1, 2009

Edited by the **Institute of Information Theories and Applications FOI ITHEA®**, Bulgaria,
in collaboration with the **V.M.Glushkov Institute of Cybernetics of NAS, Ukraine**,
and the **Institute of Mathematics and Informatics, BAS, Bulgaria.**

Publisher: **ITHEA®**
Sofia, 1000, P.O.B. 775, Bulgaria. www.ithea.org, e-mail: info@foibg.com
Printed in Bulgaria

ISSN 1313-0455 (printed) **ISSN 1313-048X (online)** **ISSN 1313-0501 (CD/DVD)**

EXPERIMENTAL SUPPORT OF SYNTACTIC COMPUTATION BASED ON SEMANTIC MERGE OF CONCEPTS

Velina Slavova, Alona Soschen

Abstract: *Linguistic theory, cognitive, information, and mathematical modeling are all useful while we attempt to achieve a better understanding of the Language Faculty (LF). This cross-disciplinary approach will eventually lead to the identification of the key principles applicable in the systems of Natural Language Processing. The present work concentrates on the syntax-semantics interface. We start from recursive definitions and application of optimization principles, and gradually develop a formal model of syntactic operations. The result – a Fibonacci-like syntactic tree – is in fact an argument-based variant of the natural language syntax. This representation (argument-centered model, ACM) is derived by a recursive calculus that generates a mode which connects arguments and expresses relations between them. The reiterative operation assigns primary role to entities as the key components of syntactic structure. We provide experimental evidence in support of the argument-based model. We also show that mental computation of syntax is influenced by the inter-conceptual relations between the images of entities in a semantic space. This work represents a further step in the formal description of the observed syntax-semantics dependencies. The assumption is made that the syntax-semantic interface is best explained as a particular operation, Merge that applies at both semantic and syntactic levels. The resulting formal description of the stages of syntactic treatment complies with the results of the experiment.*

A formal model of syntactic operations is developed starting from recursive definitions and application of optimization principles. The result – a Fibonacci-like syntactic tree – is in fact an argument-based variant of the natural language syntax.

Keywords*: natural language, mathematical modeling, cognitive modeling*

ACM Classification Keywords*: I.2 Artificial Intelligence, 1.2.0. Cognitive simulation*

Introduction

We use mathematical formalism of Generalized Nets to develop a stage-simulating model of NLP. This formal approach allows a more exact representation of information flows during the stages of processing, expressed as the transitions Z_1–Z_{29} of the Net (Slavova 2004). The analyses performed on this basis suggest that information treatment consists of the operations that use two types of Long Term Memory knowledge (syntactic and semantic) in parallel. As an example, this is the case of transition $Z_{27,}$ which expresses the stage when the system builds the syntactic structure of a sentence after its last word-form was stored in Working Memory (figure 1.). A detailed examination of the incoming information flow allows us to suggest that the procedure, running on $Z_{27,}$ must use semantic and syntactic knowledge in parallel. We assumed that **syntactic structure is better clarified** when it receives semantic justification.

For further analyses, the two types of knowledge stored in Long Term Memory were modeled by means of a database structure that shows the interconnection of syntactic rules, semantic primitives, and semantic operators (Slavova, Soschen, Immes, 2005). The assumption was that language units (word-forms) have images as semantic primitives such as “concepts”, “attributes”, “events” etc, and that grammatical rules comply with semantic operations on these primitives. This formalization of the Language as a “joint” Information System was used to study a particular language rule - secondary predication in Russian[1]. This rule was modeled by means of the formal approach described above. That led to a coherent and well-defined formal procedure and confirmed that the rule entails operations on semantic primitives.

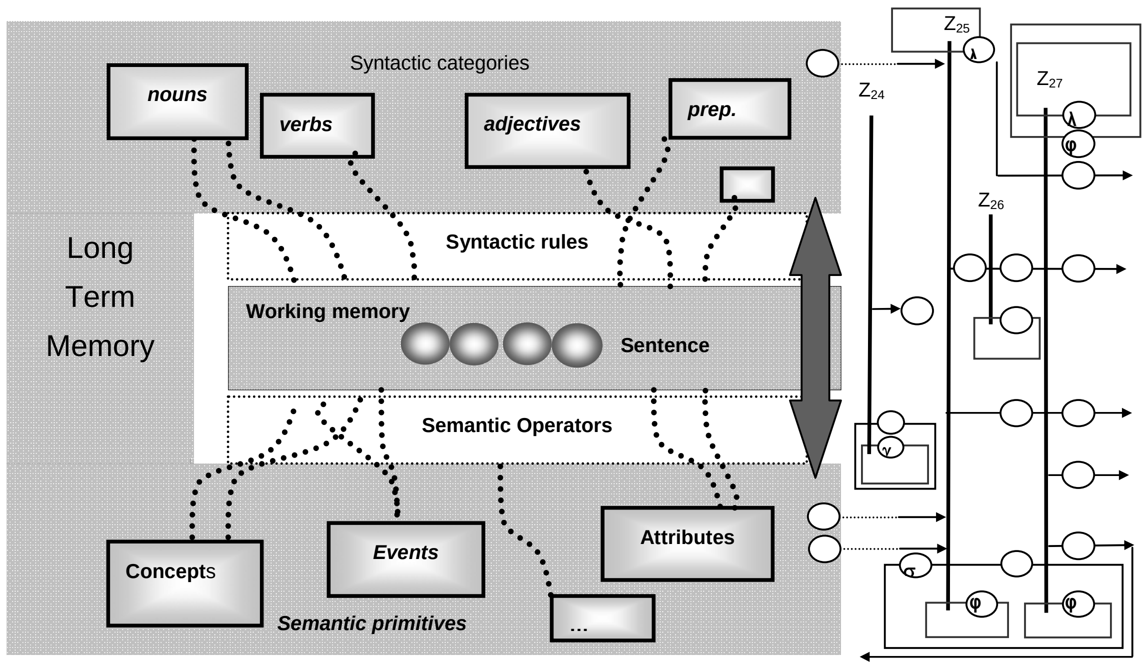

Figure 1. Information treatment of a sentence, based on language and semantics

Further efforts are put forward to obtain the proof that **semantic knowledge** and **syntax** are interrelated. The question so far is how syntax is related to operations on semantic primitives – concepts, events, attributes, etc. This is one of the most important questions in contemporary linguistics and cognitive science.

Syntax as Computation

Following one of the widely accepted linguistic theories, the key component of Faculty of Language (FL) is a computational system (narrow syntax) that generates internal representations and maps them into the conceptual-intentional interface by the (formal) semantic system (Hauser et al., 2002). There is a consensus that the core property of FL is *recursion*, which is attributed to narrow syntax. In other words, the process of mental generation of syntactic structures relies on the capacity of the human brain to perform specific operations in compliance with the principles of efficient computation. The claim in the recent theories is that this computation is based on a primitive operation that takes already constructed objects to create a new object. This basic operation, called “Merge”, provides a “language of thought”, an internal system to allow preexistent conceptual resources to construct expressions (Chomsky, 2006). Although these questions receive a lot of attention, there

[1] The linguistics theories don’t provide a consistent explanation of Secondary Predication in Russian.

are no convincing proposals yet concerning the precise type of resources on which such computation is performed in a recursive manner to build syntactic structures.

Following from the above, the study of syntactic recursion by mathematical means may provide valuable insights into the principles underlying the human language. One step in this direction was provided in Slavova and Soschen (2007). Syntactic structures, presented in the traditional sense of Chomskyan theory (Bare Phrase Structures, XP-structures), were re-defined in terms of finite recursive binary trees. The "traditional syntactic tree" does not correspond to the finite nature of a sentence; consequently, it cannot be defined recursively as a finite object. Another reason to introduce this modification is to build a structure that complies with the principles of optimization, namely with the principle of efficient growth (Soschen 2006, 2008). The tree was modified; the nodes related to syntactic role of verbs were discarded. The structure obtained in this way is a tree of Fibonacci (figure 2. a).

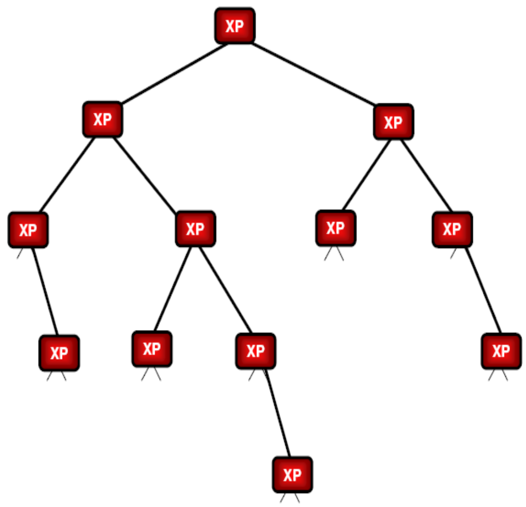

Figure. 2a.

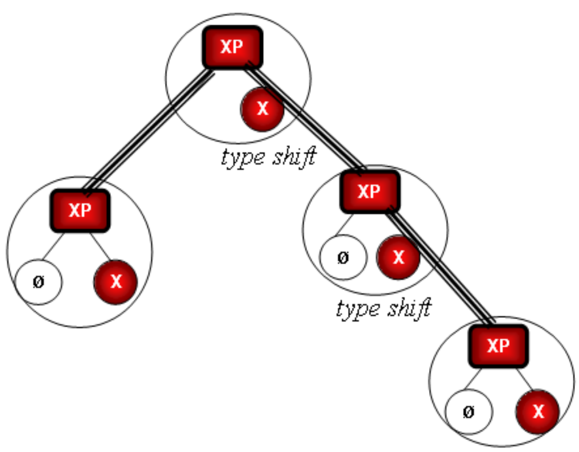

Figure. 2. b.

This tree can be seen as is an operator – it "performs" a bottom-up Merge (figure 2.b.), its nodes are the results of Merge. In the model under development, XPs are *sets*, Xs are 'unbreakable' *entities*, and Merge can be applied to two non-equivalent substances (the tree has ordered nodes). These formal transformations of the traditional tree result in a structure that incorporates two operations of fundamental importance in the syntactic model. The first is "Ø-Merge", operation that takes place at the point where Xs as initial substances form *singleton sets,* ready for further syntactic computation. The second is *type-shift,* which results in a transition from *sets* (XPs) to *entities* Xs and expresses a property of the dual mental representation of XP as either consisting of two separate elements or as an 'unbreakable' whole (part of a larger unit).

The Fibonacci-like tree shows the patterns of relating arguments (Soschen 2006, 2008). An important question is the height h of the XP Fibonacci-tree, since it refers directly to the memory, necessary for the computation. The tree is a recursive object; the same patterns of Merge are repeated at its levels. It is easy to show that merge-patterns start to reiterate when h>3 and that any tree with h>3 can perform more than one merge-pattern. We defined the tree with h=3 as the basic tree (fig. 2.b). We interpret its properties as follows: the basic tree defines the maximal number of Xs that can be merged in a procedurally unambiguous way. It could be suggested that this structure is determined in the same way as the number of nodes and relations that can be treated by the human brain within a semantically meaningful argument space. The tree represents a bare (label-free) syntactic structure that has no lexical input; what it has are the paths that connect smaller units in order to produce a larger meaningful unit. We called the tree in (fig. 2.b) "the Argument-Based Syntactic Tree".

According to the hypothesis put forward in Soschen (2005, 2006, 2008), a general rule governing efficient growth applies in syntax in such a way that minimal syntactic constituents incorporate arguments (*agent, recipient, theme*) which are related to each other. In the Fibonacci-tree model, the type of merge configuration determines the type of relation between arguments. The maximal configuration (fig. 3.d) corresponds to thematic roles *agent*, *recipient*, and *theme*. The "syntactic meaning" of the schemes in (fig. 3) corresponds to configurations offered in (Soschen 2006, 2008). These schemes represent all possible configurations and relations between arguments in the human theta-role Semantic Space. Carnie (2006) shows convincingly that the number of arguments in a thematic domain is necessarily limited to three, a fact that has not found an explanation in linguistics so far. The model under development suggests that the number of arguments is limited in a particular way in compliance with the principles of efficient growth, which are, in our terms, the principles of efficient computation as well.

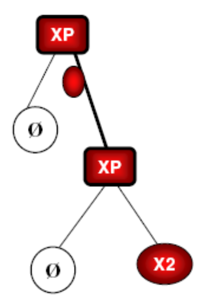

Figure 3.a. Infinite iteration: Mary, Mary…

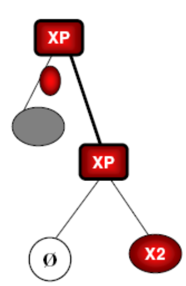

Figure 3.b. Mary in *Mary smiles*.

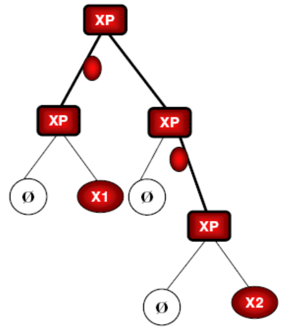

Figure 3.c. Two arguments Mary, John in Mary loves John.

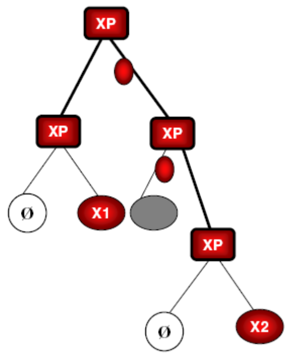

Figure 3.d. Three arguments Mary, John, apple in Mary gave John an apple.

Of importance to linguistic theory is our proposal that the argument-based model of syntax has a fundamental character. This model shows that syntax utilizes recursive calculus to connect *arguments* and express *relations* between them. The argument-based model assigns a primary syntactic role to *entities*, usually expressed as nouns. This viewpoint is in contrast with *verb*-centered models of syntax. Our efforts are focused on the experimental evidence that supports the argument-based model. The difficulty of designing an appropriate experiment is that mental computation runs on a deep (pre-linguistic) level and cannot be captured on the lexical level by a standard experiment. One possible way to extract some information about the primary mechanisms is to force the mental system to solve ambiguities on the lexical level and to analyze the system's response.

Experimental Design

Bulgarian is the only Slavic language which, during the last 10 centuries, has undergone a transition from synthetic to analytical language. Prepositions replaced case-flections, and a suffixed definite article appeared. One interesting result of the transition is that the Genitive and Dative cases are both expressed by means of the preposition *'на'* (na). "Na" has several meanings: to, of, on. Our experiment is based on the following two meanings of "Na":

1. *Of – meaning* (whose, Slavonic Genitive)

The X *на* *the* Y means “the X of the Y” i.e. “the Y’s X”, as in:

The X	Ha	The Y	
Къщата	На	Кучето	
The house	Of	The dog	The dog's house

2. To – meaning (to whom, Slavonic Dative)

Subject Verb *на* *the* Y means that the subject S acts To the Y. For transitive verbs, на assigns the syntactic role of a *Recipient*:

S Verb O		The Y (Recipient)
Той донесе стол	На	Директора
He brought a chair	To	the director

In the example above, Object is not marked with an article. Such sentences always have the meaning S-(V)-O-R (three arguments: agent, theme, and recipient).

When the Object is marked with an article, the sentence becomes:

Subject Verb *the X (*Object) *на* *the Y* .

and its second part fits the Genitive construction *the* X *на* *the* Y . In result, the available grammatical rules of the language assign to the noun Y two possible roles:

1. Subject Verb *the* X (Object) *to* *the Y (*Recipient). S-(V)-O-R, Resipient (1)

2. Subject Verb *the* X (Object) *of* *the* Y (Possessor). S-(V)-O-of-P, Possessor (2)

In such sentences, preposition *на* indicates that the noun that follows it is either Recipient (argument), or it is the object’s owner/ Possessor. The difference between these two interpretations is crucial, as the basic syntactic structure of two sentences is completely different - in the former, there are three arguments, and in the latter, there are two (corresponding respectively to the trees on fig. 3.d and 3.c). In Bulgarian, all the sentences of type:

Subject Verb *the* Object *на* *the* Y.

are ambiguous: they assign two different meanings to Y - *Recipient* and *Possessor*.

In normal listening or reading-comprehension conditions, native Bulgarian speakers interpret one of these meanings depending on the context. The sentence “Mary gave the book на the boy.” in the context “Mary entered holding a book and she saw a boy” is interpreted as “Mary gave the book to the boy.” And, in the context “The boy left his book. Mary was asked about the book.” the very same sentence is interpreted as “Mary gave the boy’s book to someone else.” Speakers of Bulgarian are never mistaken about the conveyed meaning. However, as our experiment has shown, they are not even aware of the existence of the two meanings. It appears that in the cognitive space such “*на*-sentence” acts as a Necker Cube – one may “see it” in either of the two ways. The context makes one of the meanings explicit, while the subjects are not aware of the other meaning. And, in fact, as is the case with Necker’s Cube, if one concentrates long enough on an isolated *на*–sentence, one will discern that it has two meanings.

Our goal is to study the mechanisms of mental computation of the syntactic structure of an isolated sentence, with regard to the role of the verb and the arguments.

1. If the assumption is correct that the argument-centered computation is the key to mental operations, an isolated *на*-sentence will be constructed by assigning to Y the role of Recipient.
2. The *на*-sentences are ambiguous; if the role of *entities* (nouns in this case) is primary, semantic relations between their images in the conceptual nets will influence the final result of the syntactic computation.

Experiment

In what ways an isolated *на*–sentence is interpreted? We prepared 13 examples of *на*-sentences (Table 1). Each of these sentences has an argument that conveys either of the two meanings – Recipient (Rc) vs. Possessor (Ps). All the verbs used in the test examples are transitive and allow Recipient. All the sentences can exist as complete sentences without Possessor and without Recipient. The verbs are in the past tense, Perfective form.

Table 1.

200.Ex	Иван Ivan	Продаде Sold	Къщата The house	На to/of	баща си his father
201.Ex	Мария Mary	Продаде Sold	Колата the car	На to/of	Съседката the neighbour
202.Ex	Михаил Mihail	Продаде Sold	Къщата The house	На to/of	съседа си his neighbour
203.Ex	Елена Elena	Продаде Sold	Къщата The house	На to/of	Кучето the dog
204.Ex	Анна Anna	Продаде Sold	Ябълките The apples	На to/of	Момчето the boy
211.Ex	Анна Anna	Подаде Gave	Стола the chair	На to/of	Директора the director
212.Ex	Петър Peter	Донесе Brought	Стола the chair	На to/of	Директора the director
220.Ex	Мария Mary	Показа Showed	Колата the car	На to/of	Съседката the neighbour
221.Ex	Иван Ivan	Показа Showed	Пътеката the wolk	На to/of	Баща си his father
222.Ex	Петър Peter	Показа Showed	Къщата The house	На to/of	Баща си his father
231.Ex	Кумчо Вълчо The Big Bad Wolf	Продаде Sold	Къщата The house	На to/of	Кучето the dog
232.Ex	монтьорът The fitter	Показа Showed	Колата The car	На to/of	Съседката the neighbour

We need to find out which of the two meanings of these isolated sentences is obtained FIRST, i.e. in the most natural way. That can provide information about the mechanisms of mental computation of the basic syntax.

The difficulty in designing an efficient experiment is that when asked to explain the meaning of such a sentence, subjects usually reply by repeating the very same sentence. For them, in the first moment, the sentence has only one meaning that can be put into words in one particular way only. The subjects do exactly what they were asked to do: they express the meaning by using words. Further efforts to make them reveal the meaning make them focus on the sentence for a longer period of time. As a result, they discover that the sentence has one more meaning, and they report that the sentence can mean two different things.

This difficulty was overcome in a tricky way. We used the fact that sentence structure, including word order, is exactly the same in French. The crucial difference is that the preposition *на* is translated in French as “à” (to) for the Recipient-meaning and as “de” (of) for the Possessor-meaning.

The subjects of our experiment were the students in the masters program of the Francophone Institute for Management in Sofia[1], all of them fluent speakers of French. The subjects, 62 students with different backgrounds (economists, sociologists, biologists, linguists, engineers etc.), were: native speakers of Bulgarian - 39, of Ukrainian - 6, of Rumanian – 5, of Russian - 3, of Georgian – 3, of Albanian – 3, of Macedonian – 2, and of Arabic – 1. Some of the non-native Bulgarians spoke Bulgarian fluently, some were less fluent.

The statements in Bulgarian were presented in a written form to the subjects, on small separate pieces of paper, with the only instruction “Translate into French”. It was done at the end of regular classes, under circumstances implying that “it is not something you should worry about, do it speedily”.

Each statement was presented to 10-12 different subjects. Each subject was given 2 different statements in a random manner, while the statements did not contain the same verb or the same noun. The 23 non-native Bulgarian speakers could ask the experimenter about the meaning of Bulgarian words. There were a few questions about the meaning of “монтьор” (fitter), “тапицер” (upholster) and “пътека” (path) as well as about the corresponding French-tense of the verbs (Past-perfect forms are translated with “passé composé”). There were no questions about the meaning of *на*.

The 124 written translations of the test statements were stored in a database. Table 2 contains the proportion of the Recipient- and Possessor-meanings assigned to each statement (Of% and To%).

This experimental design was successful in the sense that only 4 subjects, native Bulgarian speakers, became aware that a given sentence has 2 meanings. It is interesting that some of these subjects noticed the double meaning of one of the statements that they had to translate, but not of the other. They were asked to put down the two possible translations in the order in which the meanings came to their minds, and only the first one was taken into account for further analyses.

The results in Table 2 show that, in spite of the “Necker’s cube property” of each statement, one of its possible meanings is interpreted by the subjects more often than the other. The second observation is that for some statements the preferred interpretation is the Recipient-meaning and for others – the Possessor-meaning**.** The third observation is that these changes do not depend on the verb. For one and the same verb, the interpretation “switches” from one to the other meaning. For example, as one can see in Table 2, “Sold” appears in statements varying from 100% of Recipient-meaning, to 100 % of Possessor-meaning.

[1] Institut de la Francophonie pour l’Administration et la Gestion - Sofia

Table 2.

	Subject	Verb	Object	Ha	Y	Of%	to%	Tendency
204.Ex	Anna	Sold	The apples	To/of	the boy		**100**	Y = Recipient
202.Ex	Mihail	Sold	The house	To/of	his neighbor	29	71	Y -> Recipient
201.Ex	Mary	Sold	The car	To/of	the neighbor	30	70	Y -> Recipient
231.Ex	The Big Bad Wolf	Sold	The house	To/of	the dog	33	67	Y -> Recipient
200.Ex	Ivan	Sold	The house	To/of	his father	67	33	Y -> Possessor
203.Ex	Elena	Sold	The house	To/of	the dog	**100**		Y = Possessor
221.Ex	Ivan	showed	The path	To/of	his father		**100**	Y = Recipient
220.Ex	Mary	showed	The car	To/of	the neighbor	11	89	Y = Recipient
222.Ex	Peter	showed	The house	To/of	his father	33	67	Y -> Recipient
232.Ex	The fitter	showed	The car	To/of	the neighbor	50	50	Equivalence
211.Ex	Anna	gave	The chair	To/of	the director		**100**	Y = Recipient
212.Ex	Peter	brought	The chair	To/of	the director	13	88	Y = Recipient
233.Ex	The upholster	brought	The chair	To/of	the director	50	50	Equivalence

Based on the available experimental data (at least ten trials for each statement from different subjects), we assume that the experiment has captured some major tendencies in the interpretation of the test statements. This experiment allows us to further explore the principles of mental operations underlying interpretation of the basic syntactic argument structure. So far, a linguistic theory that would explain the observed tendencies in obtaining some particular result, “computed” by the subjects, has not been developed. Our experiment has shown that the explanation can be provided by using the argument-oriented model derived in compliance with the principles of efficient computation.

Analyses of Experimental Results

The experimental results show that the interpretation of the syntactic structure depends on entities (in this case, nouns). The verb itself does not predetermine the type of structure: either S-(V)-O-R (three arguments) or S-(V)-OofY (two arguments). Many of the contemporary linguistic theories mostly consider predicate-based and verb-centered syntactic structures. Actually, if the verb does not allow a recipient, the syntactic structure of the *на*-sentence is calculated as S-(V)-O of Y.

Suppose that mental calculus depends solely on the type of the verb. Then in the cases where the verb allows Rc, на would ALWAYS imply a S-(V)-O-R structure. But that is clearly not the case in the last four examples, given in Table 3 (where the examples are arranged by the “captured from subjects meaning”) :

As it is shown in Table 3, when the verb allows a Recipient, *на* implies preferably, but not necessarily the structure S-(V)-O-R (three arguments). The noun Y selects the Rc role in most cases. If mental operations were not dependent on the calculus which relies on the arguments as primary substances, all the statements of the experiment would be with around 50% interpretation of Y as Rc and 50% - Y as Ps.

We may suppose that the argument-centered representation of syntax is the key to syntactic analyses.

Table 3

Subject	Verb	Object	Ha	Y	of%	To%	Tendency
Anna	Sold	the apples	to/of	the boy		100	Y = Recipient
Ivan	showed	the path	to/of	His father		100	Y = Recipient
Anna	Gave	the chair	to/of	The director		100	Y = Recipient
Mary	showed	the car	to/of	the neighbor	11	89	Y = Recipient
Peter	brought	the chair	to/of	The director	13	88	Y = Recipient
Mihail	Sold	the house	to/of	his neighbor	29	71	Y -> Recipient
Mary	Sold	the car	to/of	the neighbor	30	70	Y -> Recipient
The Big Bad Wolf	Sold	the house	to/of	The dog	33	67	Y -> Recipient
Peter	showed	the house	to/of	His father	33	67	Y -> Recipient
the fitter	showed	the car	to/of	the neighbor	50	50	Equivalence
the upholster	brought	the chair	to/of	The director	50	50	Equivalence
Ivan	Sold	the house	to/of	His father	67	33	Y -> Possessor
Elena	Sold	the house	to/of	The dog	100		Y = Possessor

The next question is: if the argument S-(V)-O-R structure is calculated first, what are the reasons that lead the calculus to take another route and assign a S-(V)-O of Y structure to a similar sentence? Our assumption is that the sentence is kept in working memory (figure 1.) and that the final “solution” about basic syntactic roles is assigned to all its parts after semantic verification. If that was not true, the word order would be the key factor in the syntactic computation and the observed differences in the interpretation would not appear.

Let us analyze why the statement:

Elena Sold The house *to/of* The dog. | 100% of | Y = Possessor,

is interpreted as having S-(V)-O of Y structure. The reason for that seems very clear: the noun dog is rejected as Rc of “sold”. The noun takes upon itself the role of the owner of the house. If this is the right mechanism, it is sufficient to provide “the dog” with the possibility to be the Rc of the house, or to modify a noun: “Elena sold the house to a dog-buyer”.

The argument-centered syntactic model attests to the fact that syntactic relations depend on the relations between concepts that exist in the semantic space. In fact, as the experimental results show, it is sufficient to replace the subject noun with the one that can be related to the dog as a buyer in a fairy tale context:

The Big Bad Wolf Sold The house *to/of* The dog | To 67% | Y -> Recipient

This result indicates that mental calculus takes into consideration not only the meaning of the noun but also the relations between the nouns. Thus:

Ivan	Showed	The path	*To/of*	His father	To	100%	Y = Recipient
Ivan	Sold	The house	*To/of*	His father	67%	Of	Y -> Possessor

The three possible arguments of both sentences correspond to concepts that exclude relations such as "fathers have paths" or "sons sell houses to their fathers". Note that sentences reveal the relations between all the three of the arguments. The predominant meaning in the semantic space of the second sentence is 'fathers have houses and sons operate their father's property'.

These dependencies between the basic concepts expressed as Subject and Object are shown as two pairs of statements below:

Subject	Verb	Object	Ha	Y	of%	To%	Tendency
Mary	Showed	The car	to/of	The neighbor	11	89	Y = Recipient
The fitter	Showed	The car	to/of	The neighbor	50	50	Equivalence
Peter	Brought	the chair	to/of	The director	13	88	Y = Recipient
The upholster	Brought	the chair	to/of	The director	50	50	Equivalence

When Mary shows the car, she shows it TO the neighbor; when the fitter shows the car, there is a high probability that this is the neighbor's car. In the semantic space, fitters operate on cars, while neighbors have cars. That same tendency is observed in the second in pair (upholsters and a director's chair). Once again, argument structure is influenced by the inter-conceptual relations.

These examples provide evidence about the nature of the primary elements - participants in mental operations. It becomes clear that syntactic computation depends on the meaning of the nouns and inter-conceptual relations.

AGN Tracking

AGN is a Generalized Net model, developed for simulating the cognitive process of natural language comprehension (Slavova 2004). This formal approach allows a more exact representation of information flows during the stages of processing, expressed as the transitions Z_1—Z_{29} of the Net. AGN has been used as formalism in several studies in cognitive linguistics. The detailed examination (Slavova, Soschen (2005)) of the information flow has led to the suggestion that the procedures, running on Z_{25} and Z_{27}, must use semantic and syntactic knowledge in parallel (figure 4). Contemporary linguistic theories don't provide the necessary theoretical bases for the formalization of the process (the most part of them don't agree that syntax is related to semantics). The presented work aims to give further development of AGN by analyzing the procedures which perform simultaneously in the language and the semantic space. Transition Z_{25} (figure 4) simulates two processes, which run in parallel - the activation of the semantic space σ (the Semantic net NSet stays on l_{70}) by the message word-forms Wi and the detection of grammatically related word-chains in the sentence.

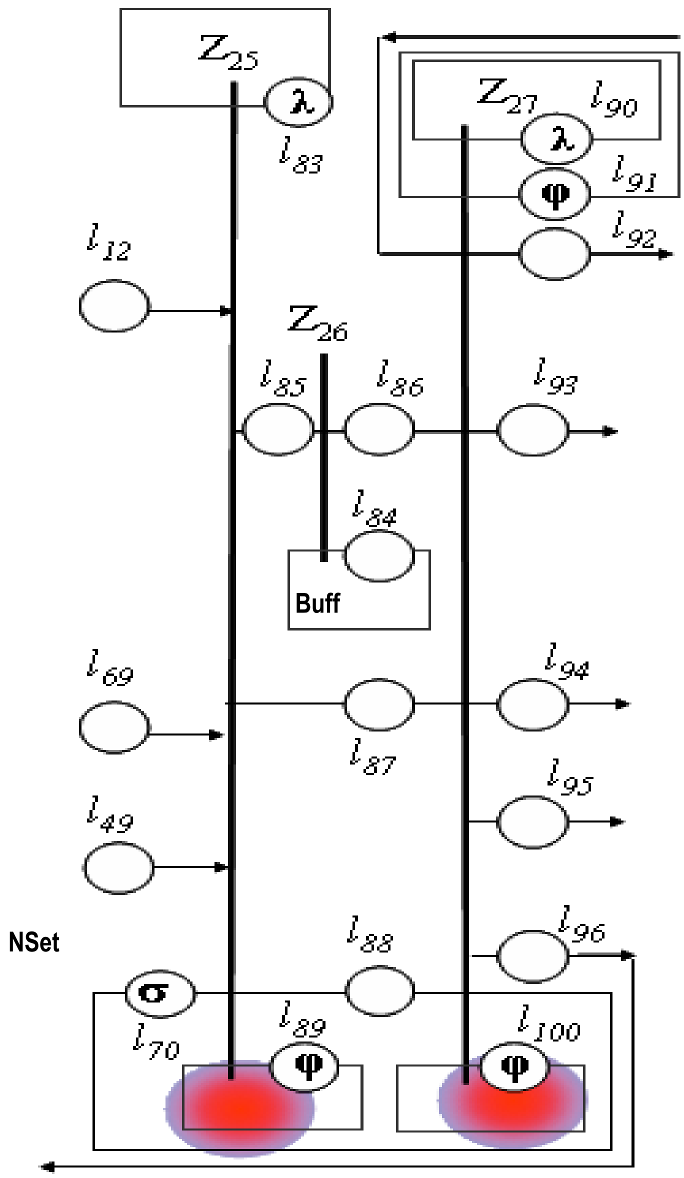

Figure 4. Transitions, based on language and semantics

It is supposed that the cognitive system first assembles a fractional representation of the sentence-meaning structure (coupled words for example).

The incoming information flow for Z_{25} is:

l_{12} – Wi+GrFtrs - word-forms Wi, with their grammatical features GrFtrs;

l_{49} - Ntct - the list of nodes of the semantic NSet, corresponding to Wi and the couples WNk: (Wi, NSetj), representing the word-forms Wi, assembled with their corresponding nodes of NSet.

l_{69}, - NtSBlist - the list of attributes of the concepts

The running on Z_{25} procedures are:

TreeBranches on I_{83} (λ-token), which obtains a partial syntactic tree:

$$\text{"ParSynStr = TreeBranches (NSet, GrFtrs)} \quad (3)$$

SemA (φ-token) on I89 which stores activation ANet in the semantic net NSet:

$$\text{ANet = SemA (Ntct, NtSBlist)} \quad (4)$$

Transition Z_{26} represents a Working Memory buffer (Buff). It stores the W_i + GrFtrs, with their corresponding semantic nodes – the couples WN_k: (W_i, $NSet_j$) and represents the "lexical memory", storing the words with their meanings.

Transition Z_{27} expresses the mental process of analyzing the entire sentence after its last word-form has been perceived. It employs from I_{88} the activation ANet of the semantic net NSet as well as its structure. The following procedures are running on Z_{27}:

λ-token - Procedure "*Parse*" on place I_{90} with *"Syntax structure discovery"*

φ-token - Procedure "*Comp*" "Comparing semantics and syntax" on I_{91}

The previous analyses (Slavova, Soschen 2005) led to the supposition that the procedure *Parse* entails a "***Merge concepts***" operator, performed in the semantic space on Z_{27}. This supposition is developed in the analysis here. The information processing on transitions Z_{25} and Z_{27} is analyzed for the statements of type:

[X] [Verb] [the Y] [*на*] [the Z].

Z_{25}: Transition Z_{25} receives from its "language" input I_{12} the word-forms W_i with their grammatical features GrFtrs, such as gender, plural, articles etc. (Table 4).

Following grammatical features, procedure TreeBranches assembles "X" with the verb "Verb" as well as the preposition Na with the noun Z which follows it.

$$\textit{TreeBranches}\ (0, \text{GrFtrs}) = \textit{ParSynStr} = \{\{X, \text{Verb}\}, \{\textit{на}, Z\}\} \quad (5)$$

Table 4.

Wi:	GrFtrs:		
X	Noun	3 pers.	Sing.
Verb	Verb	3 pers.	Sing.
Y	Noun		/article
Na	Preposition		
Z	Noun		/article

Table 5.

NSeti semantic images:
X* - the concept of X
V*- the image of the achievement (?O, R)
Y* - the concept of Y
To* Of*
Z* - the concept of Z

The words W_i have corresponding images in the semantic net NSet, given in Table 5. These nodes and their related nodes are activated by the procedure *SemA*.:

$$\text{ANet} = \{X^*, V^*, Y^*, (To^*, Of^*), Z^*\} \cap \{\{rel\ X^*\}, \{rel\ V^*\}, \{rel.\ Y^*\}, \{rel\ Z^*\}\} \quad (6)$$

Z_{27}: To the input of Z_{27} come results from:

I_{86} – the content of Buff: W_i + GrFtrs (Table 2) and WN_k: the couples WN_k: (W_i, $NSet_j$)

I_{87} – the edges of the partial tree “ParSynStr = {{X, Verb}, {*на*, Z}}”

I_{88} – The semantic net NSet with the activated nodes ANet.

Transition Z_{27} has to perform the language procedure Parse(Buff, ParSynStr) and to perform the semantic check Comp (Parse(), ANet).

Syntactic Procedure with Semantic Merge

Procedure Parse on I_{90} uses knowledge of syntactic rules. The assumption is that it runs by executing operation “Semantic Merge”, based on the information, available on the inputs of Z_{27}.

Semantic Merge (M) is further modeled as a binary operation, performed in sequential progression between the concepts X*, Y*and Z* inside of a sentence with Verb V*. The main assumption here is that Semantic Merge is in compliance with the principles of the argument-based syntactic model. The result of Semantic Merge consists in temporal semantic images μ, which stay in working memory till the end of the syntactic treatment.

The syntax structure starts to be assembled on the bases of the grammatical information on the input of Z_{27}. The edge: *{X, Verb}* entails[1] a Semantic Merge between X* and V

As supposed in the argument-based syntactic model, the subject starts first the treatment.

$$M (X^*V^*) = [X^*, V] (\underline{?O}, ?R) \tag{7}$$

The result of M consists of a couple, in which each element obtains an image μ, representing the concept in the semantic context of the other member of the couple. For example, the image μ of the concept X* within the couple [X*V] is the image of X* as Subject, performing V:

$$\mu X^* \in [X^*V] = X^* \text{ Acts } (\underline{?O}, R) \tag{8}$$

The other grammatical edge on the input is {*на*, Z}.

Preposition *на* indicates that the noun that follows it is either Recipient, or Possessor.

The final syntactic result depends on the decision upon the meaning of *на*. Following the experimental data (Table 3), the final syntactic result is not assigned depending on *на.*

As it is seen in table 3, *The Boy* is totally rejected as a Possessor of *the Apples* without any reason residing within the couple *Boy-Apples*. In fact, all the used Z* can possess Y* in a specific context. The only possible interpretation of this experimental result is that the meaning of *на* is assigned later, depending on the obtained images μ of X*, Y* and Z* within the action V*. That means that, as suggested in AGN model, the word-forms and their semantic images stay available in working memory until the end of the syntactic treatment.

---

[1] The word order in Bulgarian is flexible and the fact that the statements are in canonic word-order S-V-O is not a sufficient condition to assign to X the role of Subject.

The grammatical information brought to the input of *Parse* is not sufficient to build a syntactic structure. The treatment can continue only by applying other mechanisms.

The schemes on figures 5 and 6 represent the supposed steps of syntactic treatment with Semantic Merge for the Recipient assignment and the Possessor assignment. The images μ obtained at each step are given in order to figure out the mechanism of the treatment and to analyze it in comparison with the experimental results.

Following the experiment, the result of one of the treatments is **rejected**. The assumption is that the intermediate images μ of either the Recipient scheme (figure 5) or the Possessor scheme (figure 6) are rejected by activating the semantic relations preexisting in NSET.

The temporal semantic result of step 2 of each scheme is:

$$\mu Z^* \in [\mu X^*, Z^*] = \text{Recipient}(\mu X^*) \quad (9)$$

$$\mu Z^* = (Z^* \leftarrow Y^*) = \text{Possessor of } Y^* \quad (10)$$

The experimental results in Table 3 show that the treatment follows the Recipient scheme (fig. 5) even when there are no reasons to reject μZ* as Possessor of Y*. The explanation of this result is that **the argument-based syntactic structure is "calculated" as primary**. If there are no reasons to reject it, it is accepted as final. That confirms the assumption that the argument-based syntax has a fundamental character.

Table 6. Influence of the Possessor relation Z*←Y*

μ X*		Z*	Y*	to%
Anna	Sold	the boy	The apples	100
Mihail	Sold	His neighbor	The house	71
Ivan	Showed	His father	The path	100
Peter	Showed	His father	The house	67

However, as the experimental results in table 6 show, if in the semantic space there exists a previous knowledge about Z* as a natural Possessor of Y*, that influences the process in step 2. The preexisting relations of Possessor Z*←Y* between *father* ← *house* and *neighbor*← *house* influence the final decision which starts switching to the Possessor decision.

Further syntactic treatment must assign to the merged couple an Object (?O), because the verbs used are transitive, and the statements necessarily must have an Object.

In the Recipient scheme (fig. 5) this is made in step 3, where a Merge is performed between μX* and Y*. The result is [μX*, Y*] where Y* is the Object of μX*. In step 4 the treatment terminates by merging [μX*, Y*] and [μ X*, Z*].

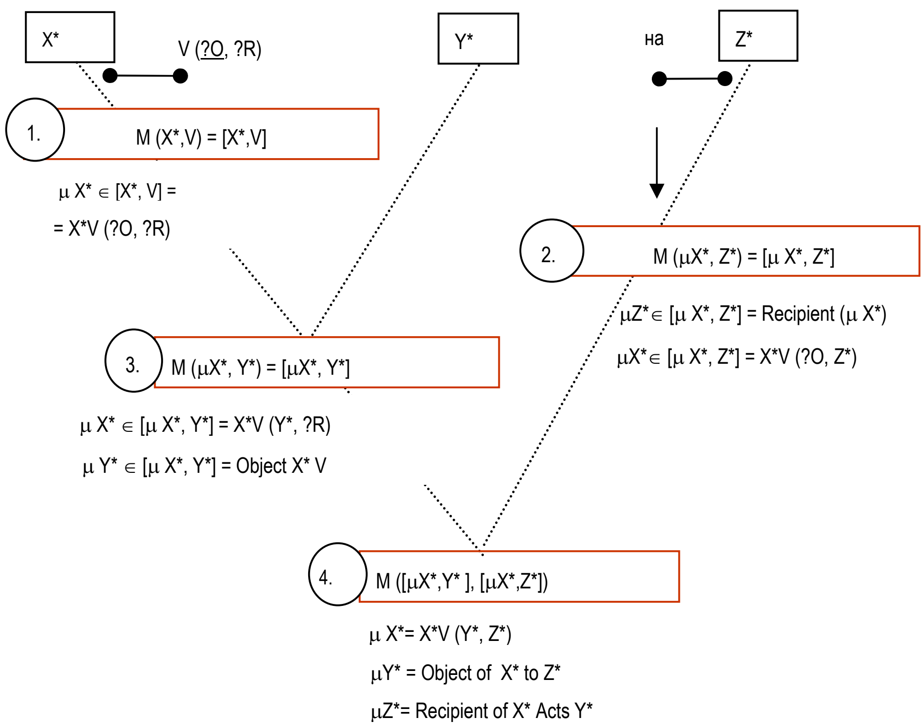

Fig. 5. Recipient scheme

In the Possessor scheme (fig. 6), Z* is assigned Possessor of Y* in step 2a. In step 2b the treatment “knows that Z* is busy” and assigns Ø to the Recipient position of μX*. In step 3 μY*=(Y*→Z*) is assigned as Object of μX*. The result [μX*, μY*] is merged in step 4 with [μX*, Ø].

The assignment of Object is influenced by the relations existing in the semantic space. The knowledge about Y* as the usual Object of actions of X* forms in the semantic space a relation X*-Y*. This relation matches the result of syntactic treatment when merging μX* and Y*. Such is the case with the couples *Fitter–Car* and *Upholster-Chair* in the examples:

The pre-existing relation X*-Y* “pushes” the treatment to Merge μX* and Y*directly, as it is in the Possessor scheme. In terms of AGN semantic activation, the concept of *the Fitter* and of *the Car* will be activated as images of Wi and as related to each other nodes. The same is with *the Upholster* and *the Chair*.

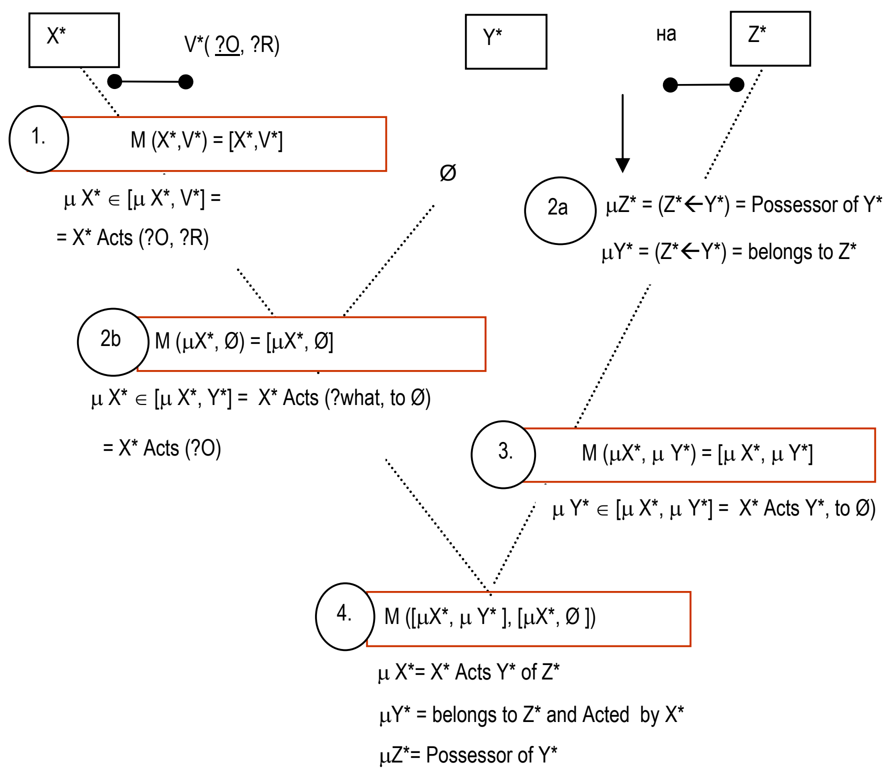

Fig. 6. Possessor scheme

The pre-existing semantic relation X*-Z* also influences the treatment. As seen in Table 7, *The Dog* is rejected as Recipient of *Elena Sold* and accepted as Recipient of *The Big Bad Wolf Sold*. *The Dog* is not rejected as Recipient of *Sold* in general. In terms of semantic activation of concepts and features, the concept of *The Big Bad Wolf* activates animals as possible Recipients of the actions. This example shows that when between X* and Z* there exists a strong semantic relation, Z* takes the (natural) role of Recipient of the actions of X*.

Table 7. Influence of the relation X*-Y*

X*	Y*			Z*	To%	Assignment
Mary	The car	Showed		the neighbor	89	Recipient
The fitter	The car	Showed		the neighbor	50	Equivalence
Peter	The chair	Brought		The director	88	Recipient
The upholster	The chair	Brought		The director	50	Equivalence

Table 8. Influence of the relation X*-Z*

μX*		Z*		Y*	Result	Assignment
Elena	Sold	The dog		The house	100% of	Possessor
The Big Bad Wolf	Sold	The dog		The house	67% to	Recipient

The existing semantic relations between all the three of the concepts X* Y* and Z* influence the syntactic treatment in the example below, where the Recipient result is rejected in the second sentence and it has a Possessor interpretation.

Table 9.

μX*		Y*	Z*	Of%	To%	Assignment
Mihail	Sold	The house	His neighbor	29	71	Recipient
Ivan	Sold	The house	His father	67	33	Possessor

The pre-existing semantic Possessor relation Z*←Y*: *Father* ← *house* pushes the treatment to the Possessor tendency. Following the result in table 6, that influence is less strong than as observed in this example. The existing strong relation X*-Z*: *Ivan-his Father* entails a Recipient tendency (table 8) and there are no reasons to reject [μX* Z*] *IvanSold-his Father.* The experimental result could be explained with the existence of a strong relation X*-Y*: *Ivan – the house* which influences the treatment (table 7) in addition of *Father*← *house*. Such a strong relation X*-Y* could exist if:

(X*-Z*) And (Z* ← Y*) ⇒ (X*-Y*)

(*Ivan - his Father*) And (*his Father* ← *House*) ⇒ (*Ivan – house*)

In the semantic space “Ivan usually operates with his father’s property”.

The general conclusion is that people construct the meaning of what has been said.

Conclusions and Future Work

The general suggestion in our approach is that interactions exist between the purely language features and some semantic fundamental which is similar for all languages.

Assumptions about how the argument structure is computed have led to the development of the argument-based model of basic syntax. The reiterative operation assigns a primary role to entities as the key components of syntactic structure.

Experimental evidence is provided in support of the argument-based model. The semantic role of entities (nouns) seems to be primary in syntax. The experimental data show that mental computation of syntax is influenced by the inter-conceptual relations between the images of entities in a semantic space. The analysis provided here is based on the assumption that the syntactic treatment includes a Merge operation between the images of the concepts. The formal description obtained of the stages of syntactic treatment corresponds to the experimental results. That supports the hypothesis that the grammatical rules entail operations on semantic primitives.

Further study will require a more precise picture of the dependencies between semantic primitives, lexical items, and syntactic rules. That will lead to an advanced modeling of the phenomenon under examination.

Bibliography

[Carnie, 2006] Carnie, Andrew, (2006) Syntax: A Generative Introduction. Blackwell.

[Chomsky, 2004] The Generative Enterprise Revisited. Mouton de Gruyter.

[Chomsky, 2006] Noam Chomsky. Biolinguistic Explorations: Design, Development, Evolution, 2006, West Hall, Bathish Auditorium, AUB.

[Hauser et al., 2002] M. Hauser, N. Chomsky and W.T. Fitch. The Faculty of Language: What is it, who has it, and how did it evolve? In: Science Vol. 298.

[Slavova, 2004] Slavova, Velina (2004) A generalized net for natural language comprehension. In: Advanced Studies in Contemporary Mathematics, vol 8, Ku-Duk Press, 131-153.

[Slavova, Soschen, 2007] Slavova, V., and A. Soschen (2007), A Fibonacci-tree model of cognitive processes underlying language faculty, in: Proc. Of 3-rd international conference in Computer Science, NBU, University of Fulda, Boston University, pp. 196-205.

[Slavova, Soschen, Immes, 2005] Slavova, V., Soschen A. and L. Immes, (2005), Information processing in a cognitive model of NLP, in: International Journal Information theories & applications, vol 12, N3, pp 157 – 166

[Slavova, Soschen, 2008] [Soschen, Slavova 2007] Slavova V., and A. Soschen (2008), Experimental support of argument-based syntactic computation, in: proc. of ITA 2008, Joint International scientific Events on Informatics, Varna 2008.

[Soschen, 2005] Soschen, Alona. 2005. Derivation by phase: Russian Applicatives. Canadian Linguistic Association Conference proceedings.

[Soschen, 2006] Soschen, Alona (2006). Natural Law and the Dynamics of Syntax (MP). Linguistics in Potsdam 25. Optimality Theory and Minimalism: a Possible Convergence? Hans Broekhius and Ralf Vogel (eds.): ZAS, Berlin.

[Soschen, 2008] Soschen, Alona (2008). On the Nature of Syntax. Bio-linguistics Journal, V.2/2.

[Soschen, Slavova, 2007] Soschen A., and V. Slavova (2007), Cognitive modeling of recursive mechanisms in syntactic processing. In: proc. of the IX international conference in Cognitive Modeling n Linguistics, Text processing and cognitive linguistics, pp 334-343

Authors' Information

Velina Slavova *- New Bulgarian University, Department of Computer Science, 21 Montevideo str., 1618 Sofia, Bulgaria, e-mail: vslavova@nbu.bg*

Major Fields of Scientific Research: Cognitive modeling in linguistics, formal models of language phenomena

Alona Soschen *- Massachusetts Institute of Technology, MIT Department of Linguistics and Philosophy, 77 Massachusetts Ave. 32-D808Cambridge, MA 02139-4307, USA, e-mail: soschen@mit.edu*

Major Fields of Scientific Research general features of biological systems present in linguistic structures