BULGARIAN ACADEMY OF SCIENCES

CYBERNETICS AND INFORMATION TECHNOLOGIES • Volume **5**, No 2

Sofia • 2005

# Exercise in Conceptualization*

*Kamenka Staykova*

*Institute of Information Technologies, 1113 Sofia*
*E-mail: staykova@iinf.bas.bg*

***Abstract:*** *This article aims at knowledge representation through conceptualization of natural language verbs. It describes a particular attempt to conceptualize the most frequent Bulgarian verbs against the conceptual model called The Globe and constructed on the base of the computationally implemented ontology GUM.*

***Keywords:*** *knowledge representation, ontologies, natural language processing, syntetic functional grammar.*

## 1. The conceptual model

### 1.1. Introduction

According to S o w a [6] "Knowledge Representation is a multidisciplinary subject that applies theories and techniques from three other fields:

1. *Logic* provides the formal structure and rules of inference.
2. *Ontology* defines the kinds of things that exist in the application domain.
3. *Computation* supports the applications that distinguish Knowledge Representation from pure philosophy.

Without logic, a knowledge representation is vague, with no criteria for determining whether statements are redundant or contradictory. Without ontology, the terms and symbols are ill-defined, confused, and confusing. And without computable models, the logic and ontology cannot be implemented in computer programs. Knowledge representation is the application of logic and ontology to the task of constructing computable models for some domain."

The definition "Ontology is an explicit specification of a shared conceptualization" G r u b e r [9] is widely accepted within the society of knowledge engineers. The scientific areas of Computational Linguistics and Natural Language Processing

* This work was supported by project IIT-010061 "Technologies of the Information Society for Knowledge Processing and Management".

have their own traditions and attitude to ontologies, mainly connected to specificity of linguistic knowledge organization. There are different understandings and approaches to construct ontology among the researchers of Computational Linguistics. The process of conceptualization could be commenced in different ways and conducted in different directions.

The exercise presented in this article is based on the following assumptions:

1) the conceptual model we aim is a linguistically motivated one, which means that all its concepts have linguistically motivated parameters. As a consequence, we should answer the questions of dependency.

2) is the conceptual model domain (and task) independent?

3) is the conceptual model language independent?

Let's consider all the concepts of an eventual model divided into the following four groups (Table 1).

Table 1

| Type | Domain Independent | Domain Dependent |
|---|---|---|
| Language Independent | General Level of Knowledge (for generic concepts and relations – time, space, causality etc.) | Conceptual Terms in Particular Domain of Knowledge |
| Language Dependent | Linguistic Knowledge, Common throughout the Domains | Conceptual Framework of Low Level Formally Presented Info/Knowledge |

The conceptual model described in this article contains domain independent concepts only and the presented exercise in conceptualization concerns Bulgarian language in particular. The ontology chosen among the number of existing linguistic ontologies[1] to give shape of the presented conceptual model is Generalized Upper Model.

## 1.2. Ontology GUM

"The Generalized Upper Model is a computationally implemented, general, task and domain independent, "linguistically motivated ontology" intended for organizing information for expression in natural language. It is a descendent of the Penman Upper Model, originally developed by Bill Mann, Christian Matthiessen and others at the USC/ISI in Los Angeles." (B a t e m a n et al., [1]) Generalized Upper Model (GUM) is the operational ontology of the Multilingual Environment for Natural Language Generation KPML.

The theoretical motivation of GUM comes from the Systemic Functional Grammar (SFG), the fundamental work of M. A. K. Halliday (H a l l i d a y [4]). SFG interprets a natural language "as a system of meanings, accompanied by forms through which the meanings could be realized." (H a l l i d a y [4, p. XIV]). Like in most linguistic theories, SFG recognizes the sentence taking 'a significant border post' [H a l l i d a y [4, XX-XXI]) in the natural human speech (spoken or written).

SFG uses "the classroom image of grammatical structure: Language is made up of sentences. A sentence consists of clauses, which consist of groups (or phrases), which consist of words, which consist of morphemes." (H a l l i d a y [4, p. 23]); but "the image" is used in completely new way – applying the strategy "to adopt the

[1] John Bateman's page **http://www.fb10.uni-bremen.de/anglistik/langpro/webspace/jb/info-pages/ontology/ontology-root.htm** offers an useful classification of ontologies.

framework of sentence, clause, group, word and morpheme as a strict hierarchy of constituents, each one being related by constituency to the next."

SFG is sensitive to the three different strands of clause meaning:

"– a clause has meaning as a message, a quantum of information;

– a clause has meaning as an exchange, a transaction between speaker and listener;

– a clause has meaning as a representation, a construal of some process in ongoing human experience." (H a l l i d a y [4, p. 23]).

The current paper concerns a formal representation of general knowledge, so the focus is on the layer *clause as a representation*. "Language enables human beings to build a mental picture of reality to make sense of what goes around them and inside them. Here again the clause plays a central role, because it embodies a general principle for modelling experience- namely, the principle that reality is made up of PROCESSES.

What is the status of a process, as set up in the grammar of the clause? A process consists, in principle, of three components:

(i) The PROCESS-itself;

(ii) PARTICIPANTS in the process;

(iii) CIRCUMSTANCES associated with the process.

This tripartite interpretation of processes is what lies behind the grammatical distinction of word classes into verbs, nouns, and the rest, a pattern that in some form or other is probably universal among human languages."(H a l l i d a y [4, 106-109]).

Different PROCESS TYPES are introduced further after thorough analysis of English speech (Table 2).

A number of implementations over Systemic Functional Grammar are created and Generalized Upper Model is a part of them. Theoretically established PROCESS TYPES are used within GUM ontology as basic concepts. The structure of GUM begins with the most general notions UM-THING and UM-RELATION and this way it is split into two hierarchies.

The top node of the first hierarchy shown in Fig. 1, UM-THING, corresponds to the most general abstract entity in the semantics of the GUM and presents a linguistically defined "Phenomena" or "Situations". There are three major subtypes of UM-THING:

- Single, "stand-alone" object or conceptual item; this is represented by the concept ELEMENT.

Table 2. Process types, their meanings and key participants (H a l l i d a y, [4, p. 143])

| Process type | Category meaning | Participants |
|---|---|---|
| Material<br>Action<br>Event | "doing"<br>"doing"<br>"happening" | Actor, Goal |
| Behavioural | "behaving" | Behaver |
| Mental<br>Perception<br>Affection<br>Cognition | "sensing"<br>"seeing"<br>"feeling"<br>"thinking" | Senser, Phenomenon |
| Verbal | "saying" | Sayer, Target |
| Relational<br>Attribution<br>Identification | "being"<br>"attributing"<br>"identifying" | <br>Carrier, Attribute<br>Identified, Identifier;Token, Value |
| Existential | "existing" | Existent |

• Configuration of elements; all elements participate in some activity or state of affairs; this is represented by the concept CONFIGURATION.

• Complex situation where various activities or CONFIGURATION are connected by some relation to form a sequence; this is represented by the concept SEQUENCE.

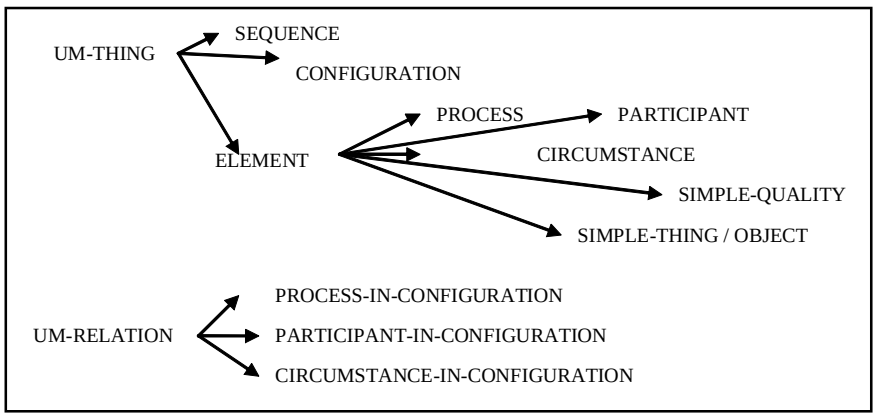


Fig. 1. Top level of the two GUM hierarchies

The following excerpt comes from the GUM documentation [Bateman at al, 1995]: “Entities classified under PROCESS can usually be expressed as verbs and are frequently the main verb in a clause; this contrasts with entities classified under CONFIGURATION which would be realized by the clause itself. “

Generally speaking, CONFIGURATIONS consist of PROCESS, PARTICIPANT(S) and CIRCUMSTANCES or, much accurately, the particular performance depends tightly on the PROCESS TYPE. The following three schemas add details to the picture of the three subtypes of CONFIGURATIONS presented by Generalized Upper Model: DOING&HAPPENING, BEING&HAVING and SAYING&SENSING. The classification discussed theoretically in Systemic-Functional Grammar is firmly followed here, although a sketchy comparison could suggest something else.

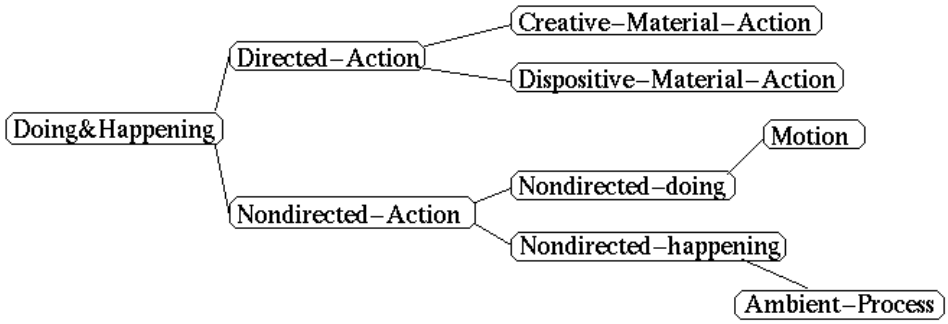


Fig. 2. DOING&HAPPENING hierarchy of GUM

“Doing and Happening” corresponds to MATERIAL PROCESS described by Halliday ( Table 2).

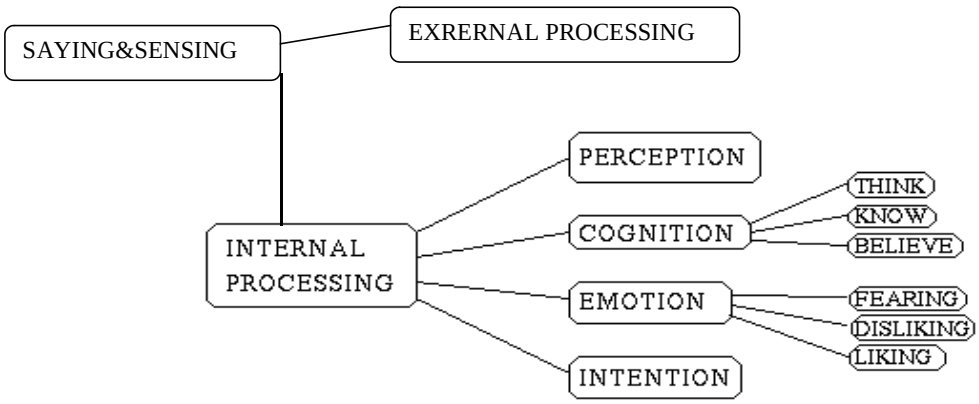


Fig. 3. SAYING&SENSING hierarchy of GUM

"Saying and Sensing" presents the three types: VERBAL PROCESS, MENTAL PROCESS and BEHAVIOURAL PROCESS (Table 2).

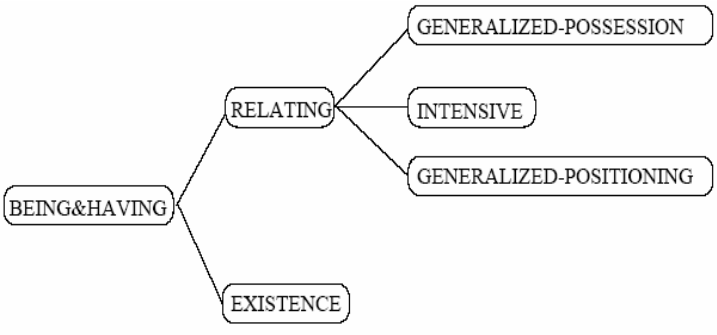


Fig. 4. BEING&HAVING hierarchy of GUM

"Being and Having" covers Halliday's RELATIONAL PROCESS and EXISTENTIAL PROCESS (Table 2).

### 1.3. Systemic-functional analysis of Bulgarian

The ontology GUM is "generalized", because its conceptual structure is verified for English, German and Italian (B a t e m a n et al. [1]). The question whether the construction of this ontology is appropriate for other natural languages raises a very interesting discussion concerning multilinguality.

The theoretical analysis of Bulgarian shows that GUM supports the backbones of the different CONFIGURATION types in Bulgarian:

- in Bulgarian the clause takes the same position on the border between smaller constructive elements and bigger logically connected chunks of text;
- terms PROCESS, PARTICIPANTS and CIRCUMSTANCES could be used with their Systemic-Functional meaning, because there is a high level of similarity of the way they are realized in Bulgarian and in English in the clauses; the PROCESS is

usually realized by verbal group, PARTICIPANTS – by nominal group and CIRCUMSTANCES – by adverbial group or prepositional phrase.

• we claim that the same PROCESS TYPES described in English are presented in Bulgarian (S t a y k o v a, P e n c h e v, [8]) and almost the same evidences could be used to demonstrate their existence and differences among them in Bulgarian speech as they are in English.

The conclusion is that the most upper part of the Generalized Upper Model could be used as representation of "linguistic knowledge common throughout the domains" in Bulgarian. In general, we could build a Bulgarian conceptual model following the theoretical Systemic Functional principles in analysis and presenting examples in Bulgarian analogous to English ones. Table 3 below shows the key concepts supported by examples, which are extracted from Hallidey's book or from the GUM documentation.

### 1.4. The Metaphorical Image of Globe

Halliday claims that the PROCESS TYPES are ordered and they form a circle as is showed on the Figure below, taken from (H a l l i d a y [4, p. 108]). The author even mentions, that "…more accurately still, they (the types of processes) could be shown to form a sphere, but that becomes too complex metaphor to handle" (H a l l i d a y [4, p. 107]). The author of this paper attempts to create a formal representation of "the metaphor of sphere" called The Globe.

The exact meaning of the used terms is given bellow to avoid possible misunderstandings and confusions. PROCESS is Halliday's PROCESS-itself and represents an action, something happening. It is a basic concept of ontology GUM and is taken to the Globe Model. CONFIGURATION is a concept defined in GUM and represents a particular PROCESS together with its PARTICIPANTS and CIRCUMSTANCES. VERB is a classical grammatical term, it is a kind of word and could be conjugated within a particular clause, for example, in tense, person and so on. PROCESS and CONFIGURATION are notions which carry meaning, sense, but they could be "technically" realized and made visible in natural language through words, PROCESS, in particular, is expressed mainly through VERBS. Logically, the area of semantic models will be sensitive to terms PROCESS, CONFIGURATION, PARTICIPANTS, CIRCUMSTENCES and so on. It is a fact, that we have nothing as real as natural speech, with its words, among which VERBS, and there should be a kind of connection between them and an ontological construction. An investigation in the direction VERBS → *ontological concepts* is presented in the second part of the article. The description of The Globe conceptual model follows bellow.

The Globe could be figured as a 3-dimensional construction organized round a point called the Center of the Globe. Now, let's consider three rays which emanate from the Center and lie in one and the same plane. It probably will be easier if we imagine that each ray radiates light in different color: red, green, blue. Let's associate the Center of the Globe with the concept CONFIGURATION taken from the GUM ontology. Let's orientate the three groups of GUM CONFIGURATIONS – DOING&HAPPENING, BEING&HAVING, SAYING&SENSING to the three rays (Tabl. 3). Near the Center the colors of the rays merge and become paler and we can call the Center "the white point". Far away from it the colors are brighter and distinguishable from each other.

Table 3. Key CONFIGURATIONS in English and Bulgarian obtained by theoretical analysis

| The Different Area Meanings = CONFIGURATION TYPES | | | | | Process in English | Process in Bulgarian |
|---|---|---|---|---|---|---|
| DOING&HAPPENING | Directed Actions | Dispositive Material Actions | | | DO<br>GIVE<br>KILL<br>WAKE<br>EAT | ПРАВЯ, ВЪРША<br>ДА(-М-ВАМ)<br>УБИ(-Я-ВАМ)<br>СЪБУ(-ДЯ-ЖДАМ)<br>ИЗЯ(-М-ЖДАМ) |
| | | Creative Material Actions | | | MAKE<br>CREATE<br>WRITE | ИЗРАБОТ(-Я-ВАМ)<br>СЪЗДА(-М-ВАМ)<br>НАПИ(-ША-СВАМ) |
| | Nondirected Actions | Nondirected Doings | | | STIR<br>GO<br>TREMBLE<br>FAINT | ДВИЖА СЕ<br>ОТИ(-ДА-ВАМ)<br>ТРЕПЕРЯ<br>ПРИПАД(-НА-АМ) |
| | | Nondirected Happenings | | | HAPPEN<br>COLLAPSE<br>WAKE | СЛУЧ(-И-ВА) СЕ<br>СРУТ(-Я-ВАМ) СЕ<br>СЪБУ(-ДЯ-ЖДАМ) СЕ |
| | | Ambient Processes | | | RAINING SUNNING | ВАЛИ<br>ПЕЧЕ |
| BEING & HAVING | Existence | | | | BE EXIST | СЪМ<br>СЪЩЕСТВУВАМ |
| | Relating | In-ten-sive | Identity | | BE | СЪМ |
| | | | Symbolization | | BE ? | СЪМ |
| | | | Ascription | Class | BE | СЪМ |
| | | | | Property | BE | СЪМ |
| | | | | Quantity | BE | СЪМ |
| | | Generalized Positioning | | | BE | СЪМ |
| | | Generalized Possession | Part-Whole | | HAVE ? | ИМАМ |
| | | | Ownership | | HAVE<br>POSSESS | ИМАМ<br>ПРИТЕЖАВАМ |
| | | | Name-of | | HAVE | ИМАМ |
| | | | Generalized Role Relation | | HAVE | ИМАМ |
| | | | Ascription Inverse | | HAVE | ИМАМ |
| SAYING & SENSING | Internal Processing | Perception | | | PERCEIVE | ВЪЗПРИЕМ(-А-АМ) |
| | | Cognition | | | THINK | МИСЛЯ |
| | | Emotion | | | SENSE FEEL | ЧУВСТВАМ |
| | | Intention | | | WANT | ИСКАМ, ЖЕЛАЯ |
| | | Mental Active | | | CONVINCE | УБЕ(-ДЯ -ЖДАВАМ) |
| | External Processing | Proper Verbals | Addressing | | TELL<br>ASK | УВЕДОМЯ(-ВАМ)<br>ПИТАМ |
| | | | Message Transfer | | EXPLAIN<br>SAY | ОБЯСНЯ(-ВАМ)<br>КА(-ЖА-ЗВАМ) |
| | | | Communicative Attitude | | COMPLAIN<br><br>TALK | ОПЛА(-ЧА-КВАМ)<br>СЕ<br>ГОВОРЯ |
| | | Behavioral Verbals | | | SOB | РИДАЯ |

If we consider a fixed Radius of the Globe we could examine the Globe's surface. If we face the source of the (let's say) red light from the surface at the point where the red ray comes to it, we would stay in a spot of dark red color. There will be hues at the areas around, where the dark red fuses with dark green and dark blue. If we plumb a bit from the surface to the Center of the Globe, we'll see much lighter and not so bright colors around.

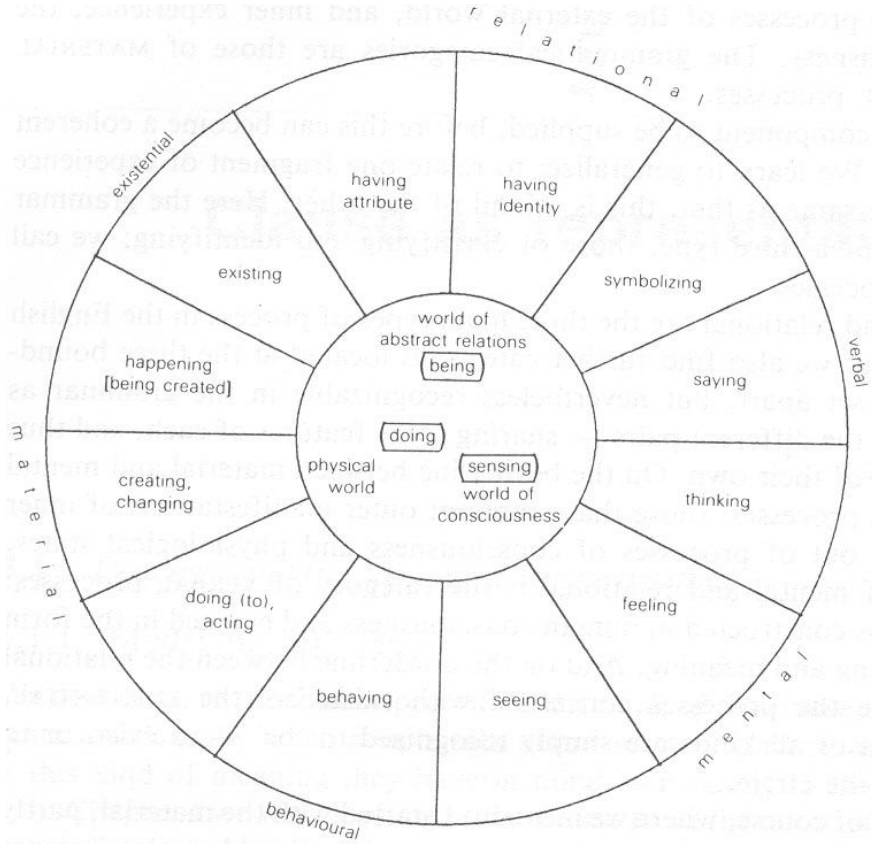


Fig. 5. The grammar of experience: types of process in English

The Globe has a spherical shape and consists of points, each corresponding to particular CONFIGURATION. Some points disposed closer to each other form areas with their specific characteristics. The areas of different kinds (Dispositive Material Actions, Creative Material Actions, Nondirected Doings and so on) should be distinguishable observing the Globe model.

CONFIGURATIONS expressing very typical meaning of particular CONFIGURATION TYPE should be disposed near the corresponding ray and marked with the most saturated color of this ray. This happens, because the points of the rays possess all particular characteristics (indications) of a given CONFIGURATION TYPE.

CONFIGURATIONS of "additional" PROCESS TYPES, sharing characteristics of two of the "main" PROCESS TYPES and introduced by Halliday as BEHAVIOURAL, VERBAL, EXISTENTIAL, should be placed as points between

the areas of main CONFIGURATION TYPES. They should be marked in fused colours as violet, orange and so on. Inside the Globe, near the Centre, the CONFIGURATION-points present very general meanings and appear in light colours.

The main idea behind the Globe is the understanding that the more particular CONFIGURATION-points are bounded by more abstract ones and even by concepts which couldn't be expressed through a CONFIGURATION of natural language.

## 2. Exercise in conceptualization

The complex ontology we make endeavour to present is a Globe model in Bulgarian, which contains the most important CONFIGURATIONS concerning domain independent concepts. This is a very ambitious task, because a suitable metric system in 3-dimentional space is required to present each CONFIGURATION-point by coordinates in a frame of reference, where one of the axes stays for "generalization". The substantial information needed is the correspondence of CONFIGURATION meaning to its position within the Globe model, which raises some very difficult problems, for example, the estimation how general is a particular abstracted CONFIGURATION; how far is its meaning from the meanings (CONFIGURATION-points) already presented within the Globe, and so on. The most challenging unknown function is that of the correspondence meaning_presented_by_the model ↔ meaning_realization_via_ the_speech. This is the problem, which stimulates wide range of research in one or another form, evolving all kinds of methods- from linguistic ones, to statistical, mathematical and so on.

The hypothesis followed here is that the most important configurations in respect to the Globe model structure are the configurations obtained by the presented above not quite thorough theoretical analysis of Bulgarian. The established simple exercise of conceptualization aims to present some evidences supporting this hypothesis.

Our "tentative offer" for most important CONFIGURATIONS is the set of most frequent CONFIGURATIONS in Bulgarian. Semantics is usually the most inaccessible layer in natural language analysis. There are not many resources in Bulgarian concerning semantics of speech chunks. Some results of BulTreeBank project are resources in usable computational form, but there is not much semantic information presented for each sentence (tree) in the corpus. Concerning SFL analysis, "meaning" is a function of three layers of analysis for each chunk of speech and even if the main process in a clause belongs to one particular PROCESS TYPE there is no guarantee, that this particular meaning is used in the particular case not modified by intonation, rhythm or metaphorical modes of expression.

Back to the conditions of our exercise, we could only weaken the "tentative offer" for most important CONFIGURATIONS and begin with processing the most frequent Bulgarian verbs. The author finds it logically correct to process first the most frequent verbs, having in mind that the most frequently used verb forms should be those of auxiliary verbs, which help in forming complex tenses. Without any additional information than the number of occurrences, we could only assume that the most frequently used meanings of a particular verb are listed among all of the meanings given by the Bulgarian Interpreting Dictionary. In fact, we will use all of the meanings of a particular verb, because the first results of the exercise will give us better orientation if more CONFIGURATION-points are presented in the model. It

will be simply a first step in performing the exercise, which could be repeated with higher precision, but it is possible a rarely used CONFIGURATION-point to appear in the model now. The resources used are the only available frequency list of Bulgarian words driven by the corpus of BulTreeBank project **http://www.bultreebank.org/ Resources.html** and the Bulgarian Interpreting Dictionary (A n d r e i c h i n et al., [7]).

Formulation of the Exercise: The main purpose of the exercise is to represent the senses expressed by the most frequent Bulgarian verbs as CONFIGURATIONS-points within the conceptual Globe model. The desired result is the picture of dispersion of the points-CONFIGURATIONS corresponding to the first verbs in the list.

The exercise is organized in two steps. The first one aims to find out which are the most frequent Bulgarian verbs. The second step is specification of the relations between the verb's senses and CONFIGURATIONS of the Globe; each verb is presented as a set of points-CONFIGURATIONS in the Globe model, where each point corresponds to a particular sense of the verb.

**First step.** The frequency list of Bulgarian words available from BulTreeBank project web page consists of 100 000 tokens and the number of their occurrences within the BulTreeBank corpus. The tokens in the frequency list are ordered from those with the highest number of occurrences to those with smaller number of occurrences. First 500 verb forms are sorted out of all kind of tokens in the list.

The problem to solve here is whether or not the token under consideration is a verb form. The resulting list consists of those tokens only, which could be used as verb forms in a clause. The ambiguity of some words is often encountered here; for example, the token **"става"** could be used as verb **ставам (get up)** in 3 prs., sing., present simple or as the noun **става** (joint, articulation); the token **"бели"** could be used as verb **беля (bark, peel)** in 3 prs., sing., present simple or as the noun **беля (trouble, mischief)** in plural, or as the adjective **бял (white)** in plural and so on. A simple rule is followed here to resolve the ambiguity: the number of occurrences of the ambiguous token is divided to equal parts among all possible variants of its usage as part of speech.

Now the infinitive verb form of each verb form from the list has to be determined. In Bulgarian verbs are fixed in their infinitive form when they are conjugated in 1 prs., sing., present simple. A problem arises in the cases where one and the same token could be retrieved from different verbs, for example, the token **"бил"** could present the verb form of the verb **съм (to be)** in 3 prs., sing., past tense or the form of the verb **бия (beat)** in 3 prs., sing., past tense. Again, some very simple rules are followed. All the possibilities are taken in account when there are ambiguous verb forms and the number of occurrences of the ambiguous verb form is divided to equal parts among all possible verbs in infinitive form from which the token could be retrieved.

We could order our list now by infinitive verbs and it is visible that often one and the same verb has been used in different verb forms. We can associate each infinitive form with the sum of the occurrences of all its forms. This way we get to 279 verbs out of the first 500 most frequent verb forms.

We could take as given, that the aspect of Bulgarian verbs does not change the meaning of the verb, so concerning following conceptualization, it's good to put in one and the same article the occurring verbs in their different aspects (perfect and imperfect). When the occurrences of perfect and imperfect variants of verb are summed together, the final list has 237 entries presenting a couple: verb (in perfect aspect, in

Table 4. First 500 verb forms are sorted out of all kind of tokens in the list

| | | |
|---|---|---|
| съм /to be/ | 2 414 948 | е, са, беше, бе, бъде, бяха, бъдат, съм, бил, били, сме, било, била, сте, бях, бяхме, бъда, бъдем, бъдете |
| мога /can/ | 287 051 | може, могат, мога, можеше, можем, можеш, могъл, можете, могли, можеха, могло, могла, можех, можа |
| имам /to have/ | 279 929 | има, имат, имаше, имаме, имам, имало, имате, имал, имаш, имали, имаха имах |
| нямам /to have not/ | 169 919 | няма, нямаше, нямат, нямам, нямаме, нямало |
| трябва /must/ | 164 047 | трябва, трябваше, трябвало |
| кажа, казвам /to say/ | 139 618 | каже, кажа, казах, казал, казаха, кажеш, кажете, каза, казва, казват, казвам, казваше |
| бия /to beat/ | 105 082 | би, бил, били, било, била, бих, биха |
| предложа, предлагам /to suggest/ | 83 701 | предложи, предлагат, предлага |
| зная /to know/ | 66 986 | знае, знам, знаят, знаеше, знаеш, знаете, знаем, зная, знаех |
| стана, ставам /to become/ | 65 671 | стане, станало, станат, станаха |
| искам /to want/ | 60 358 | иска, искам, искат, искаше, искаме, искаш, искал, искате, исках |
| започна, започвам /to begin/ | 57 875 | започна, започне, започнаха започнат, започнал, започнали |
| направя /to do/ | 53 446 | направи, направят, направил, направим, направя, направиха, направили |
| остана, оставам /to remain/ | 53 161 | остана, остане, останали, останат, останаха, останало |
| видя, виждам /to see/… | 52 363 | видя, види, видях, видим, видят, видял, вижда, виждам, виж, виждаше, виждал, виждат |

imperfect aspect or both) and a number of occurrences of forms of this verb in the corpus. The top of the list is shown in Table 4.

The first step of the exercise is completed.

**Second Step.** The second and substantial step of the experiment is the conceptualization of top 10 verbs against the CONFIGURATIONS of the Globe conceptual model, described in the first part of the article. The Bulgarian Interpreting Dictionary (A n d r e i c h i n et al., [7]) is used, where all the meanings of a particular verb were checked up. Below are given the results.

**съм (to be):** This is the leader with far more occurrences then the second verb in the resultant frequency list (2 414 948 vs 287 051). Although much of these occurrences are in its role of auxiliary verb it remains the most frequent verb. From functional point of view, this is because **съм (to be)** is used to realize several types of CONFIGURATIONS: Existence, Relating-Intensive-Identity, Relating-Intensive-Symbolization, Relating-Intensive-Ascription, Relating-Generalized-Positioning. All these possible CONFIGURATIONS are marked in the Table 5.

**мога (can):** auxiliary verb; has no reflection within the presented Figure of the Globe.

**имам (have):** The impersonal form **има** is most frequently used (158 464 occurrences vs 121 465 occurrences of all other forms), which suggests that the meanings **има, съществува (there is, exists)** and **има, намира се някъде (there is somewhere)** are widely used. These meanings correspond to the CONFIGURATIONS Existence and Relating-Generalized-Positioning respectively. First meaning of the verb within the Bulgarian Interpreting Dictionary is **possess** and leads to a bunch of CONFIGURATIONS under Generalized-Possession (Table 5).

**нямам (have not):** The form **няма (has not)** is used almost 3 times more then all the other forms of the verb (128 435 vs 41 484). It happens mainly because the form **няма** is quite often an auxiliary verb, when negative simple future is expressed: **няма да (will not)**, but also the meanings **няма (don't exist)** and **няма (there is not)** are possible. The main meaning of the verb according to the Bulgarian Interpreting

Dictionary is **нямам (possess not),** which is realized by CONFIGURATIONS under Generalized-Possession (Table 3).

**трябва (must, have to):** auxiliary verb; it is not depicted in the figure of the Globe.

**кажа, казвам (say):** It is probably the most typical verb to build a verbal CONFIGURATION. It could be chosen as a label of *ExternalProcessing-ProperVerbals area of* CONFIGURATIONS. Its different meanings could be expressed by CONFIGURATIONS under ProperVerbals-Addressing, ProperVerbals-MessageTransfer and ProperVerbals-CommunicativeAttitude except one, **казвам се (to be called)**, which belongs to Relating-Intensive-Symbolization.

**бия (beat):** Presence of this verb in the list of top 10 most frequent verbs is definitely a result of a very simple procedure of disambiguation applied during the experiment. The rule is to associate each of the pretenders with even parts of the number of occurrences of ambiguous verb form. In Bulgarian most of the past forms of the verb **бия (beat)** coincide with some of the past forms of the verb **съм (to be):** *бил, били, било, била, etc*. or with the forms of particle **бих (would):** *би, бил, били, бих, биха, etc*. This way the frequency score of the verb **бия (beat)** becomes much more compatible than the likely score of its substantially used verb forms. For the sake of completeness the meanings of the verb and their correspondent CONFIGURATIONS are given in Table 5. Most of the meanings of **бия (beat)** listed in Bulgarian Interpreting Dictionary could be realized by Dispositive-Material-Action CONFIGURATION, the meanings **бия, пулсирам (pulsate)** and **бия се (fight)** could be realized by CONFIGURATIONS of Nondirected-Doings.

**предложа, предлагам (offer, suggest):** The CONFIGURATIONS, which express the meanings of this verb belong to different areas: material Nondirected-Doings for **предложа, предлагам стоки (offer goods)** and verbal Message-Transfer for the meaning *suggest something to be discussed* as in **предложа, предлагам план (offer, suggest a plan).** A notion corresponding to both meanings could be depicted nearer to the Center of the Globe. The precursors area is within the Directed-Material-Actions CONFIGURATION **дам, давам (give)** with its typical structure involving three participants.

**зная (know):** This is the verb which realizes one of the key CONFIGURATIONS (think, know, believe) in Cognition area.

**стана, ставам (happen, become, get up):** When used in its impersonalized variant the verb leads to CONFIGURATIONS of Nondirected-Happenings: **стане, става произшествие (an accident happens)** or to Ambient-Processes, with CONFIGURATION'S structure resembling Relating-Intensive-PropertyAscription: **стане, става светло (dawning), стане, става хладно (it's getting cold, it's turning chilly)**. When the verb is used personally, an analogical disperse could be mentioned. On one hand, there is a meaning of movement as in **стана, ставам (get up)** which related CONFIGURATIONS are those of Nondirected-Doings; on the other hand, some changes in main participant properties could be expressed by **стана, ставам стар, зелен (get old, turn green),** or the change of status as in **стана, ставам учител (become a teacher),** which are Nondirected-Doings.

All the CONFIGURATIONS mentioned in the above short analysis, are recorded in Table 5 by their PROCESSES. The key CONFIGURATIONS from Table 3 are given in grey for the purposes of comparison.

As a conclusion the following fact could be underlined: the most frequent verbs in our list form such CONFIGURATIONS of the Globe model, most of which are

Table 5. CONFIGURATIONS in Bulgarian and English obtained from the exercise

| The Different Area Meanings = CONFIGURATION TYPES | | | | | Process in English | Process in Bulgarian |
|---|---|---|---|---|---|---|
| **DOING&HAPPENING** | Directed Actions | Dispositive Material Actions | | | DO<br>GIVE<br>KILL<br>BEAT<br>WAKE<br>EAT<br>CHANGE | ПРАВЯ, ВЪРША<br>ДА(-М-ВАМ)<br>УБИ(-Я-ВАМ)<br>БИЯ<br>СЪБУ(-ДЯ-ЖДАМ)<br>ИЗЯ(-М-ЖДАМ)<br>ПРОМЕН(-Я-ЯМ) |
| | | Creative Material Actions | | | MAKE<br>CREATE<br>WRITE | ИЗРАБОТ(-Я-ВАМ)<br>СЪЗДА(-М-ВАМ)<br>НАПИ(-ША-СВАМ) |
| | Nondirected Actions | Nondirected Doings | | | BEAT, PULSATE<br>FIGHT<br>GET UP<br>GO<br>OFFER<br>BECOME, GET, TURN | БИЯ<br>БИЯ СЕ<br>СТА(-НА-ВАМ)<br>ОТИ(-ДА-ВАМ)<br>ПРЕДЛ(-ОЖА-АГАМ)<br>СТА(-НА-ВАМ) |
| | | Nondirected Happenings | | | HAPPENS<br>HAPPEN<br>COLLAPSE<br>WAKE | СТАНЕ-СТАВА<br>СЛУЧ(-И-ВА) СЕ<br>СРУТ(-Я-ВАМ) СЕ<br>СЪБУ(-ДЯ-ЖДАМ) СЕ |
| | | Ambient Processes | | | GETTING, TURN-ING<br>RAINING SUNNING | СТАНЕ-СТАВА<br>ВАЛИ<br>ПЕЧЕ |
| **BEING & HAVING** | Existence | | | | BE<br>THERE IS<br>THERE IS NOT EXIST | СЪМ<br>ИМА<br>НЯМА СЪЩЕСТВУВАМ |
| | Relating | In-ten-sive | Identity | | BE | СЪМ |
| | | | Symbolization | | BE<br>BE CALLED | СЪМ<br>КАЗВАМ СЕ |
| | | | Ascription | Class | BE | СЪМ |
| | | | | Property | BE | СЪМ |
| | | | | Quantity | BE | СЪМ |
| | | Generalized Positioning | | | BE<br>THERE IS<br>THERE IS NOT | СЪМ<br>ИМА<br>НЯМА |
| | | Generalized Possession | Part-Whole | | HAVE<br>HAVE NOT | ИМАМ<br>НЯМАМ |
| | | | Ownership | | HAVE<br>HAVE NOT<br>POSSESS | ИМАМ<br>НЯМАМ<br>ПРИТЕЖАВАМ |
| | | | Name-of | | HAVE<br>HAVE NOT | ИМАМ<br>НЯМАМ |
| | | | Generalized Role Relation | | HAVE<br>HAVE NOT | ИМАМ<br>НЯМАМ |
| | | | Ascription Inverse | | HAVE<br>HAVE NOT | ИМАМ<br>НЯМАМ |
| **SAYING & SENSING** | Internal Processing | Perception | | | SENSE<br>PERCEIVE | УСЕЩАМ<br>ВЪЗПРИЕМ(-А-АМ) |
| | | Cognition | | | THINK<br>KNOW | МИСЛЯ<br>ЗНАЯ |
| | | Emotion | | | SENSE<br>FEEL | ЧУВСТВАМ<br>ИЗПИТВАМ |
| | | Intention | | | WANT | ИСКАМ, ЖЕЛАЯ |
| | | Mental Active | | | CONVINCE | УБЕ(-ДЯ -ЖДАВАМ) |
| | External Processing | Proper Verbals | Addressing | | TELL<br>TELL<br>ASK | КА(-ЖА-ЗВАМ)<br>УВЕДОМЯ(-ВАМ)<br>ПИТАМ |
| | | | Message Transfer | | SAY<br>SUGGEST, PROPOSE<br>EXPLAIN | КА(-ЖА-ЗВАМ)<br>ПРЕДЛ(-ОЖА-АГАМ)<br>ОБЯСНЯ(-ВАМ) |
| | | | Communicative Attitude | | SPEAK<br>COMPLAIN<br>TALK | КА(-ЖА-ЗВАМ)<br>ОПЛА(-ЧА-КВАМ) СЕ<br>ГОВОРЯ |
| | | Behavioral Verbals | | | SOB | РИДАЯ |

already outlined by the theoretical analysis and possess key roles and relatively deep positions in the Globe. Such CONFIGURATIONS are:

- СЪМ (BE) in Existence, Relating-Intensive and Relating – GeneralizedPositioning;
- ИМА, НЯМА (THERE IS, THERE IS NOT) in Existence and Relating – GeneralizedPositioning;
- ИМАМ, НЯМАМ (HAVE, HAVE NOT) in Relating – GeneralizedPossession;
- КА(-ЖА-ЗВАМ) ( SAY ) in Proper-Verbals;
- ПРЕДЛ(-ОЖА-АГАМ) (SUGGEST, PROPOSE) in Message Transfer;
- ЗНАЯ (KNOW) in Cognition;
- СТАНЕ-СТАВА in Nondirected Happenings.

The CONFIGURATIONS, which don't belong to the above group teach us how to manage the depth of the Globe. We need to derive and show explicitly their predecessors in order to find out and show their sense.

The extract of verbs presented here is very small. One possible variant to build a rich conceptual Model could be to proceed the same way and conceptualize the following verbs in resulting frequency list. We wouldn't choose this variant, because it will bring to the conceptual model many points-CONFIGURATIONS, which could not be used in the corpus at all, coming from exhaustive picture of all possible senses of a particular verb. The author prefers to stick more tightly to the senses really used in the corpus. Possible repetition of the exercise in conceptualization could take in account the following observations and conclusions:

- a PoS Tagger will be used;
- some information from BulTreeBank could be used;
- Bulgarian WordNet could be used at the level of conceptualization;
- the "beat" problem should be resolved by decision concerning auxiliary uses versus substantial uses of "to be";
- the question which are the most frequently used verbs could be answered in different way using a different frequency list, and so on.

The exercise will be repeated, because the results for now don't have the strength of evidence that the Globe construction is reliable as representation of domain independent knowledge for Bulgarian. Most important achievement is the vision of conceptual model and the procedure to extract its CONFIGURATION-points, which contain linguistic information in addition to the semantic information. Each notion of the Globe has explicit semantic meaning, connected to its position within the Globe. All positions of the points are theoretically grounded by SFG.

The Globe model is going to be presented in the form of computationally usable knowledge. All the results of the exercise are available at the webpage of the Institute of Information Technologies **http://www.iinf.bas.bg** .

## References

1. Bateman, J. A., R. Henschel, F. Rinaldi. Generalized Upper Model 2.0: Documentation. Darmstadt, Germany, 1995.
**http://www.fb10.uni-bremen.de/anglistik/langpro/webspace/jb/gum/index.htm**

2. B a t e m a n, J. A., M. Z o c k. The B-to-Z of Natural Language Generation Systems: An Almost Complete List. France, 2001.
**http://www.purl.org/net/nlg-list** (Continually updated).
3. G r u b e r, T. R. Ontolingua: A mechanism to support portable ontologies. Report KSL, Stanford University, 1992, 61-66.
4. H a l l i d a y, M. A. K. An Introduction to Functional Grammar. Second Edition. London, Edward Arnold, 1994.
5. S i m o v, K., P. O s e n o v a, S. K o l k o v s k a, E. B a l a b a n o v a, D. D o i k o f f. Language resources and tools for the creation of a Bulgarian treebank. – In: Proc. of the Workshop on Balkan Language Resources and Tools 2003, Greece.
6. S o w a, J., F. S o w a. Knowledge Representation: Logical, Philosophical, and Computational Foundations. Pacific Grove, CA, Brooks Cole Publishing Co., 2000.
7. A n d r e i c h i n, L., L. G e o r g i e v, S. I l c h e v, N. K o s t o v, I. L e k o v, S. S t o y k o v, Z. T o d o d r o v. Bulgarian Interpreting Vocabulary. Sofia, Nauka i izkustvo, 1955 (in Bulgarian).
8. S t a y k o v a, K., I. P e n c h e v. Systemic-functional gramma and Bulgarian language. – Bulgarian Language J., Publication of the Insitute of Bulgarian Language, BAS, Vol. **4-5**, 5-24 (in Bulgarian).