

Instructions and recommendations in preparation of a dissertation abstract/summary in English

The title page has to contain the following information:

- author's name;
- title of the dissertation;
- name of the supervisor/consultant (in case of a PhD thesis).

The second page of the abstract has to contain the following information:

- date of the scientific session when the thesis has been approved to be defended;
- date, time and place of the public defense;
- scientific jury;
- brief description of the contents of the dissertation – page numbers, tables, figures, appendices, number of references; if the author prefers he may include also the number of publications the dissertation is based on, H-index of these publications, and journals they are published in.

An abstract template is given on the next pages.



**BULGARIAN ACADEMY OF SCIENCES
INSTITUTE OF INFORMATION AND COMMUNICATION
TECHNOLOGIES**

Milena Nikolova Slavcheva

**SEMANTIC DESCRIPTORS OF THE REFLEXIVE FORMS
OF VERB STRUCTURES IN CONTEMPORARY BULGARIAN,
FRENCH AND HUNGARIAN**

ABSTRACT OF PhD THESIS

Supervisor: Assoc. Prof. Georgi Gargov

Approved by Supervising Committee:

Prof. Ivan Kasabov
Prof. Galia Angelova
Assoc. Prof. Kiril Simov
Assoc. Prof. Maria Stambolieva
Assoc. Prof. Elena Paskaleva

The PhD thesis was discussed and allowed to be defended during an extended session of the Department of Linguistic Modelling at IICT-BAS, which had been held on May 10, 2011.

The defense of the PhD thesis had been held on October 25, 2011 at 11:00 am in Room 218, Block 25A, IICT-BAS.

The full volume of the dissertation is 210 pages. It consists of an introduction and five chapters (p. 1-161). It includes also an application (p. 176-210). The list of references contains 145 titles (p. 162-175). The text of the dissertation includes 37 tables and 84 figures. These lists are given at the beginning of the dissertation.

Keywords: lexical semantics, verb-centered structures, eventivity frames, reflexive verbs, cognitive approach, lexical-functional grammar, conceptual structures, semantic descriptors, object-oriented system design, Unified Modeling Language, Unified Eventivity Representation

Introduction

In this thesis, the subject of research is the semantics of the reflexive forms of verb structures in Bulgarian and their semantic equivalents in French and in Hungarian. The study is in the field of computational linguistics and language technologies and is related to the application of an object-oriented approach to the representation of the linguistic units. The aim of the research work is the construction of formal semantic descriptions of the verb units as a component of an overall infrastructure for the representation of verb semantics in a multilingual setting.

The above formulated goal is achieved by fulfilling the following tasks:

1. Definition of classes of Bulgarian verbs as the source units in the comparison with French and Hungarian verbs.
2. Definition of packages of modeling elements necessary for the conceptual level description of the selected language objects.
3. Extension of the underlying model with new modeling elements necessary for the semantic description of the particular language data.
4. Systematization of the language data and construction of the diagrammatic semantic descriptors.

The semantic models of the language units conform to the theory and formalism of the *Unified Eventivity Representation* (UER)) (Schalley 2004). The UER approach to verbal semantics is a cognitive one, that is, the meaning of verbs is closely related to "concepts of events and similar entities in the mind" (Schalley 2004, p.1). UER is a graphical formalism representing in diagrams the conceptual structure of events, which consists of: 1) static modeling elements, characterising the event participants and the association relations among them; 2) dynamic modeling elements describing the participants' behaviour and their interaction with one another. The UER modeling apparatus is an application, adaptation and extension of the *Unified Modeling Language* (UML) - an international standard for graphical representation, modeling and design of object-oriented systems in the field of computer science (OMG 2001), (Fowler 2004). The product of the research work is *SemInVeSt* (*Semantically Interpreted Verb-centered Structures*) - a knowledge base of the semantics of verbs in Bulgarian and their equivalents in French and in Hungarian.

1. Methodology

The cognitive aspect of the linguistic analysis is determined by general guiding principles relating perception, thinking and language, that is, relating conceptual structures to the

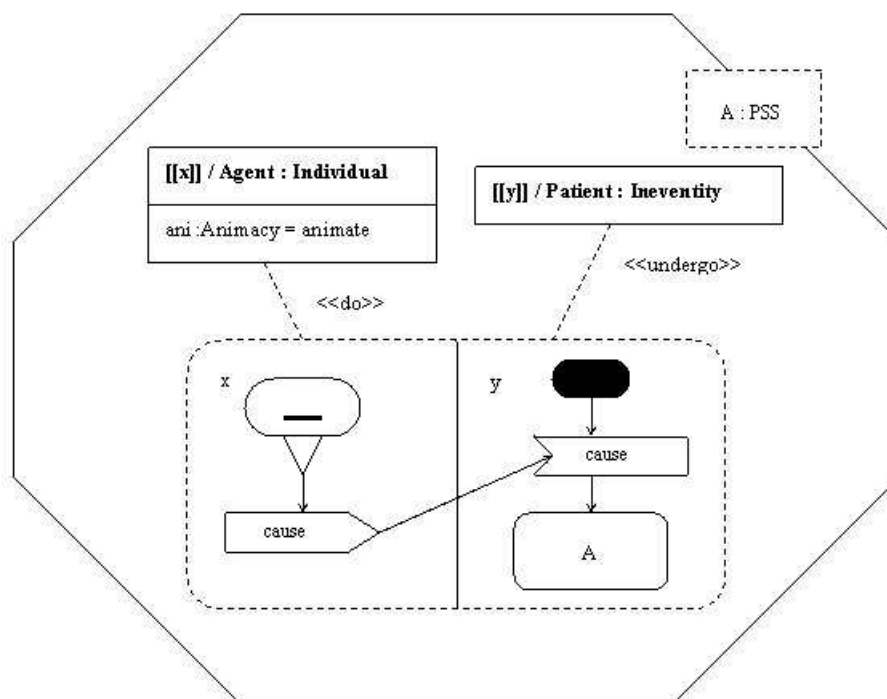
meanings and forms of the language units. In fact cognitive linguistics comprises a wide spectrum of research activities and can be more appropriately defined as an "initiative" or "movement" (i.e., "the cognitive enterprise"), rather than a uniform, strict theory (Evans et al. 2008). The general theoretical principles are based on the assumption that language reflects "patterns of thought that is why the studying of language is connected to the exploration of "patterns of conceptualization" (Evans and Green 2006). Language codes and transmits thoughts via symbols - language units, which can be parts of a word (morphemes), words, or combinations of words (phrases). Symbols consist of a form and an associated with it meaning (form-meaning pairing), that is, they are symbolic assemblies. Meaning, as a component of the symbolic assembly, is related to mental images, concepts. Concepts are evoked by percepts, which are the result of man's interaction with the surrounding world - they represent man's susceptibility to the properties of reality objects (Lakoff 1987). Percepts generate mental images - the recreated reality in the human mind.

Cognitive grammar does not make a strict division among the different levels of linguistic representation, that is, the lexical, syntactic, semantic or pragmatic level. Without rejecting completely the traditional distinction of the linguistic levels, cognitive grammar conceives the lexical and the grammatical elements as forming "a continuum of symbolic structures". That is, the grammar of a given language is a loosely defined inventory of linguistic units - entities with different phonological structure and associated with it semantic structure.

The semantic descriptions are linked to morphosyntactic knowledge about the verb lexemes. One of the sources of morphosyntactic information about the reflexive forms of verbs in Bulgarian is the Large Lexical Database of Bulgarian (Paskaleva et al. 1993), (Paskaleva 2007) and especially my work on the features, characterising the verb lexemes in a paradigmatic aspect (Slavcheva 2003a). The semantic descriptors in *SemInVeSt* are built on the basis of the semantic analysis of verb lexemes in a subset of the Large Lexical Database, that is, the Short Electronic Grammatical Lexicon of Bulgarian (Paskaleva 2003). The lexicon contains 10216 lexemes and is constructed on the basis of word frequency lists extracted from text corpora (5 million words). The Bulgarian verbs in *SemInVeSt* are excerpted from the 2465 verbs in the Short Lexicon and complemented with some verbs from the Large Lexical Database. At present *SemInVeSt* contains the semantic descriptors of 330 verb units in Bulgarian and their equivalents in French and in Hungarian.

Another source of linguistic knowledge about the reflexive verbs in Bulgarian is related to my work on the development and application of grammars for partial syntactic automatic analysis. The grammars are used for marking text chunks, which represent the so called verb complex in a syntagmatic aspect, that is a string of auxiliary verbs, pronominal clitics, particles and a full-fledged verb (Simov et al. 2002; Slavcheva 2002, 2003b,c).

The modeling of the verbal semantics in a multilingual setting is based also on underlying principles of Functional Grammar and its functional-semantic categories so as to define the relevant classes of verb units in the three languages involved. The work described in this thesis is based particularly on the theory of Functional Grammar developed by the Sankt-Petersburg linguistic circle, which has been applied in typological and contrastive studies of a multitude of languages (Bondarko 1991, 2005). The present classification of verbs is compared also with other comparative and typological studies within the framework of



Фиг. 1. Transitive EVENTITY FRAME - TEMPLATE

Functional Grammar like (Boteva 2000), (Penchev 2007) and (Guentchéva and Rivière 2007).

As pointed out above, the UER theory and formalism is applied for building the semantic descriptors of the verb units. The UER uses diagrams to represent the conceptual structure of events. The diagrams are combinations of graphical elements and linear text constructions. In Figure 1 a generic diagram of the meaning of a transitive verb is represented, accompanied by a brief discussion of the UER descriptive devices.

The central concept of eventivity is represented by an EVENTITY FRAME, whose graphical notation is an octagon containing the rest of the graphical elements denoting the eventivity components. The EVENTITY FRAME consists of a *dynamic core* and a *static periphery*. The dynamic core depicts the dynamic aspect of the conceptual structure of the eventivity and is graphically notated by a dashed-outline rectangle with rounded corners. The static periphery describes the participants, their properties and relations. The participants are represented by PARTICIPANT CLASSES. In Figure 1 the two rectangles in the upper part of the octagon denote the prominent participants, whose PARTICIPANT ROLES are specified as **Agent** and **Patient**, and whose PARTICIPANT TYPES are **Individual** and **Ineventity** respectively. The PARTICIPANT TYPES refer to ontological categories (?). In the lower rectangle of the **Agent** compartment the eligible participant is specified as **animate** - that is the value of the ATTRIBUTE named *ani* which is of the data type **Animacy**. The PARTICIPATE ASSOCIATIONS (relating the PARTICIPANT CLASSES to the dynamic core, notated by a dashed line) are specified via STEREOTYPES (<<do>> and <<undergo>>) as

the two prototypical semantic roles (i.e., macroroles, role archetypes) *proto-agent* (*actor*) and *proto-patient* (*undergoer*).

The dynamic core (the dashed-outline rectangle) is divided into SWIMLANES via the solid vertical line separating the dashed-outline rectangle with rounded corners. The SWIMLANES contain the descriptions of the participants' behaviour depicted by state-machines. In Figure 1 the left SWIMLANE belongs to the participant denoted by the x variable, and its state-machine is interpreted as follows. At some point in time the **Agent**, being in an unspecified ACTIVE SIMPLE STATE (ASS), sends a **cause-SIGNAL**, notated by the solid arrow crossing the borderline and entering the right SWIMLANE belonging to the participant denoted by the y variable. The **cause-SIGNAL** triggers a STATE TRANSITION of the **Patient** participant from a totally unspecified source STATE to a parameterized target STATE which in this case is generalized as being a PASSIVE SIMPLE STATE (PSS). The dashed-outline rectangle in the upper right corner of the octagon indicates that the EVENTITY FRAME is a TEMPLATE, which has a parameter to be bound in a specified EVENTITY FRAME.

2. Language objects described

In Bulgarian the language units described are character strings, which are a combination of a full-fledged verb form and the reflexive clitics *ce* or *cu*. Verb structures, containing the full-form of the reflexive personal pronoun *ceбе cu*, are also described. *Ceбе cu* is an unambiguous marker of prototypical, inherent reflexive meaning. Verb structures that contain the dative full-form pronoun *на ceбе cu* are also considered.

The French verb units, studied here, are a combination of a full-fledged verb form and a dependent personal pronoun (fr. pronom conjoint), which only in third person singular or plural is reflexive, that is *se*. The rest of its forms coincide with those of the personal pronouns and alter in person and number: *me*, *te*, *nous*, *vous* (*Je me promène* (I am taking a walk), *Tu te promènes* (You are taking a walk), *Elle se promène* (She is taking a walk), *Nous nous promenons* (We are taking a walk), *Vous vous promenez* (You are taking a walk), *Elles se promènent* (They are taking a walk)).

The subject of analysis in Hungarian are verb units, which contain a reflexive suffix from the following set: *-ód(ik)*, *-őd(ik)*; *-kod(ik)*, *-ked(ik)*, *-köd(ik)*; *-kód(ik)*, *-kőd(ik)*; *-koz(ik)*, *-kez(ik)*, *-köz(ik)*; *-kóz(ik)*, *-köz(ik)*; *-óz(ik)*, *-őz(ik)*; *-ódz(ik)*, *-ődz(ik)*; *-oz(ik)*, *-ez(ik)*, *-öz(ik)*; *-edz(ik)*; *-ul*, *-ül*. Some other suffixes can also add a reflexive connotation to the meaning of the verbs they attach to: the deverbal *-ad*, *-ed* and the denominal *-z*, *-l*, *-ll*. In Hungarian reflexive verb structures are also formed with the reflexive pronoun *maga*, which alters in person and number: *magam* - 1 p. sg., *magad* - 2 p. sg., *maga* - 3 p. sg., *magunk* - 1 p. pl., *magatok* - 2 p. pl., *maguk* - 3 p. pl.

The issue of reflexive verb structures is significant due to the opposition with non-reflexives. The *correlation* between reflexive and non-reflexive verb forms is the first classificational parameter. The existence or non-existence of a non-reflexive correlative of a given reflexive verb determines its further analysis and class membership. The values of the *correlation* parameter are: 1) *correlative(+)* for reflexive verb units, which have

non-reflexive counterparts (for instance, bg. *абонирам* (subscribe s.o.) / *абонирам се* (subscribe o.s.), *вдъхновявам* (inspire s.o.) / *вдъхновявам се* (be / get inspired); fr. *abaisser* (let down, drop) / *s'abaisser* (descend, go down), *libérer* (free, liberate) / *se libérer* (free, liberate o.s.); hu. *bezár* (close, shut (trans.)) / *bezárkózik* (close (intrans.), be shut), *visszahúz* (withdraw (trans.), take back) / *visszahúzódik* (withdraw (intrans.), retreat)); 2) *correlative(-)* for reflexive verb units, which do not have non-reflexive counterparts (for instance, bg. *стара се* (endeavour), *сърсявам се* (be afraid), *усмихвам се* (smile); fr. *s'efforcer* (endeavour), *s'agenouiller* (kneel), *se moquer* (mock); hu. *bánkódik* (grieve), *kínlődik* (suffer)).

The verb units classified as *correlative(+)* are numerous and are subdivided into the following classes: *inherent reflexive*, *reciprocal*, *motive*, *deaccusative*, *absolutive*, *anticausative* and *passive*.

3. Conceptual structures and modeling elements

A detailed description of the relevant for the current investigation UER modeling elements is provided in an Appendix to the thesis. In Chapter 3 the fundamental concepts are represented as well as the new modeling elements which I define as an extension to the UER model. The introduction of new elements is necessary for the construction of the semantic descriptors related to the selected language data.

Eventivity

According to the cognitive approach, the semantic representation of a given verb equals to the representation of the concept of the *eventivity*, which the verb encodes. The eventivity is a conceptual unit in the human cognition "that includes within its boundary a continuous correlation between at least some portion of its identifying qualitative domain and some portion of the so-conceived temporal continuum - that is, of the progression of time. Such a correlation may rest on a primitive phenomenological experience that can be characterized as *dynamism* - a fundamental property or principle of activeness in the world" (Talmy 2000, c.215). The model of the conceptual unit of eventivity is the EVENTIVITY FRAME.

Participant roles

One of the specifications of the PARTICIPANT CLASS is the PARTICIPANT ROLE, which is interpreted as a reference to a semantic role fulfilled by instances of the PARTICIPANT CLASS within the eventivity. Although the PARTICIPANT ROLE is used in the familiar to the linguists sense, it has its peculiarities within the UER formalism and in the structure of the semantic descriptors. The PARTICIPANT ROLES are related to *role* CLASSES, which determine the prototypical semantic roles. The *role* CLASS differs from the "normal" CLASSES because it is an ATTRIBUTE cluster, which determines a concept of a role, not a set of possible OBJECTS - instances of the CLASS. In this sense the *role* CLASSES denote conceptual configurations consisting of relational semantic features, that is, features characterizing the relation between the participants and the dynamic core of the eventivity.

The UER model suggests a set of PARTICIPANT ROLES, but it is not fixed - the user can adapt and define roles depending on the concrete task. In order to construct the

SemInVeSt descriptors, I define a system of PARTICIPANT ROLES, which conform to the UER principles, but are adjusted so as to satisfy the needs of the current semantic descriptions. Using the UER model of role hierarchies, three abstract *role* CLASSES and their subclasses are defined. The first *role* CLASS is predefined in the UER as *Instigator* with subclasses **Agent** and **Effector**. By analogy with this hierarchy I introduce the abstract *role* CLASS *Instigated*, whose elements are **Patient** and **Theme**. The abstract role *Instigated* generalizes the role of a participant who is acted upon in some way, no matter whether as a result he undergoes a change of his state or merely is localized or moved. I introduce a third role hierarchy - the abstract *role* CLASS *Medium* with subclasses **Experiencer** and **Effected**. *Medium* is a generalization of the role of a participant who occupies a middle position between a prototypical **Agent** and a prototypical **Patient**. Therefore in some eventities an **Experiencer** participant can fulfill the macro-role of an **Actor** and in others - the macro-role of an **Undergoer**. In general, the abstract role *Medium* and its elements **Experiencer** and **Effected** are part of the descriptors of eventities perceived as "happening, triggered by themselves".

Semantic relations between the participants

I use the predefined in UER CONSTRAINT **potentiallyReflexive** for modeling verb units with inherent reflexive meaning. The CONSTRAINT means that there exist PARTICIPANT OBJECTS that can be simultaneously instances of all (in practice of two) PARTICIPANT CLASSES, which are in the scope of the CONSTRAINT, that is, one diagram is used to describe the meaning of the reflexive, as well as the non-reflexive variant of a given verb.

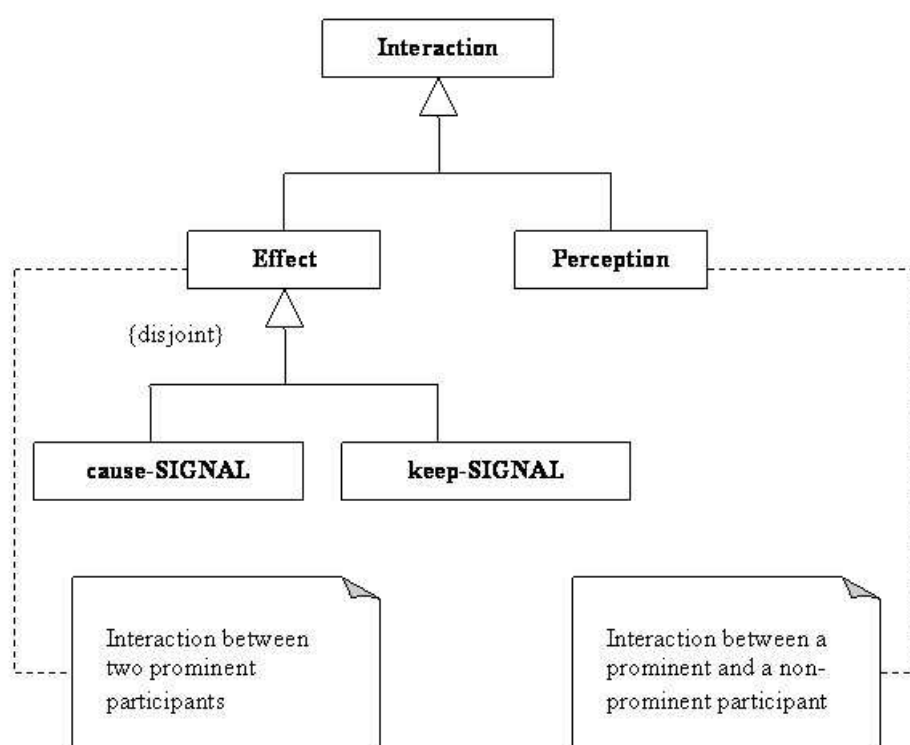
In the reciprocal verb structure (e.g., *измъчваме се (едни друг)* (torture one another)) two prominent participants interact in such a way that each one of them, performing one and the same action, affects the other participant, and, at the same time, is affected by the other participant - the interaction is symmetrical. In order to depict this type of interaction, I define a new CONSTRAINT - **Reciprocal**. It means that the PARTICIPANT OBJECTS are simultaneously instances of both PARTICIPANT CLASSES of the EVENTITY FRAME.

Eventity types and packages of modeling elements

The reflexive verb structures are modeled by two main types of EVENTITY FRAMES: 1) EVENTITY FRAMES with two participants; 2) EVENTITY FRAMES with one participant. The interaction between the two participants or the behaviour of the single participant determine the further subtyping of the eventities.

Figure 2 provides the taxonomy of the *two-participant* eventity type. In a *two-participant* eventity there are two types of interaction between the participants: 1) *effect* is an interaction between two prominent participants, where the first participant makes an impact on the second participant; 2) *perception* is an interaction between a prominent participant and a non-prominent participant, where the first participant has, generally speaking, some inner experience in connection with the second participant.

The *effect* type is subdivided into two mutually excluded subtypes (indicated by the disjoint CONSTRAINT). The first subtype of *effect* is defined as follows: a given participant makes an impact on another participant in such a way that the second participant undergoes a change of its state. The impact, as well as the change of state is taken



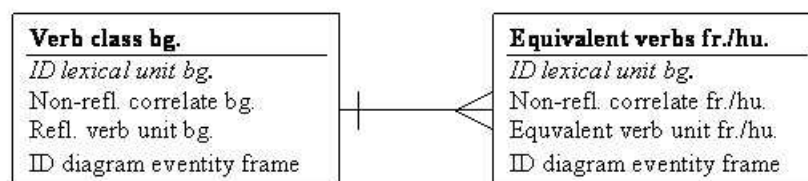
Фиг. 2. Eventity types of two-participant *interaction*

in a general sense: physical impact and triggered state transition (bg. *изяждам*, fr. *manger*, hu. *megesz* (eat up); bg. *изсушавам*, fr. *sécher*, hu. *kiszárit* (dry up)), including movement, change of place (bg. *поставям*, fr. *placer*, hu. *helyez* (put, place); bg. *продвиждам*, fr. *faire avancer*, hu. *mozgat* (move); bg. *насочвам*, fr. *diriger*, hu. *irányít* (direct)); psychical impact and state transition (bg. *убеждавам*, fr. *convaincre*, *persuader*, hu. *meggyőz*, *rábeszél* (convince, persuade); bg. *информирам*, fr. *informer qqn*, hu. *tájékoztat*, *informál* (inform); bg. *вдъхновявам*, fr. *inspirer*, *enthousiasmer*, *enflammer*, hu. *fellelkesít* (inspire)); abstract impact and change of state, related to different social activities (bg. *абонирам*, fr. *abonner*, hu. *előfizet vkinek vmire* (subscribe s.o.); bg. *прюбуцавам*, fr. *rallier*, *faire adhérer*, hu. *bevont* (join, involve)).

The second subtype of the *effect* eventity type is defined as follows: a given participant makes an impact on another participant in such a way that change of state of the second participant is not conceptualized. What is conceptualized is the first participant's pursuit of keeping unchanged the state of the second participant. This type of effect is a new one and extends the classification of conceptual structures proposed in (Schalley 2004). I also add the corresponding modeling elements to the UER metamodel. In the metamodel proposed in (Schalley 2004) a single SIGNAL is defined, that is, the **cause-SIGNAL**, which obligatorily triggers a STATE TRANSITION of the second prominent participant. However, the application of the UER modeling machinery to real-world language data shows that a considerable number of eventities are conceptualized where there are again two prominent participants, but the first participant acts in such a way as to preserve, or just not change the state that the second participant is in (bg. *пазя*, fr. *garder*, *protéger*, *préserver*, hu. *őriz* (guard, protect, keep); bg. *съхранявам*, fr. *préserver*, *conserver*, hu. *megőriz*, *tartósít* (preserve, keep); bg. *задръжам*, fr. *retenir*, hu. *marasztal*, *tartóztat*, *feltartóztat* (retain, keep, hold); bg. *държа*, fr. *tenir*, hu. *tart* (keep)). A conceptual structure of such a type necessitates the introduction of another type of SIGNAL - that is the **keep-SIGNAL**.

In the interaction between two participants, defined as *perception*, the eventity conceptualizes as inner experience of the first participant related to the second participant, for instance, a given participant perceives another participant (bg. *виждам*, fr. *voir*, hu. *lát* (see); bg. *чувам*, fr. *entendre*, hu. *hall* (hear); bg. *нунам*, fr. *toucher*, *tâter*, hu. *érint*, *tapogat* (touch, feel); bg. *броя*, fr. *compter*, hu. *számlál*, *tekint* (count, consider); bg. *наблюдавам*, fr. *observer*, *surveiller*, hu. *megfigyel* (watch, observe); bg. *меря*, fr. *mesurer*, hu. *mér* (measure)), or feels something towards another participant (bg. *обичам*, fr. *aimer*, hu. *szeret* (love); bg. *мразя*, fr. *haïr*, hu. *gyűlöl* (hate)), or expresses attitude towards another participant (bg. *обвинявам*, fr. *accuser*, hu. *vádol* (accuse); bg. *доказвам*, fr. *prouver*, hu. *bizonyít*, (prove); bg. *коря*, fr. *blâmer*, hu. *hibáztat*, *okol* (blame)). In the conceptual structure of the *perception* type there is one prominent participant, that is, the perceiver, and one non-prominent participant - the perceived.

In the *one-participant* eventities, the participant is either in a given STATE or undergoes a STATE TRANSITION and there are variants of the participant's STATE-machine depending on the presence or absence of a STATE TRANSITION, as well as on the types of source and target STATES.



Фиг. 3. One-to-many relation between the data tables

4. Semantically interpreted verb structures

Chapter 4 presents *SemInVeSt* as a knowledge base of the semantics of verb-centered structures in Bulgarian, French and Hungarian. The working copy of the *SemInVeSt* reflexive-verb-component is stored in a relational database. The data are represented in two types of tables: parent and child ones. There are eight parent tables containing Bulgarian verbs, which correspond to the eight verb classes defined in Chapter 2: *inherent reflexive*, *reciprocal*, *motive*, *absolutive*, *deaccusative*, *anticausative*, *passive*, *correlative(-)*. Each parent table contains a field of a primary key, a field of the non-reflexive counterparts of the reflexive Bulgarian verbs (with the exception of the *correlative(-)* verbs), a field of the reflexive Bulgarian verbs and a field of identifiers of EVENTIVITY FRAME TEMPLATE diagrams, which provide the semantic description of the lexical objects in the database.

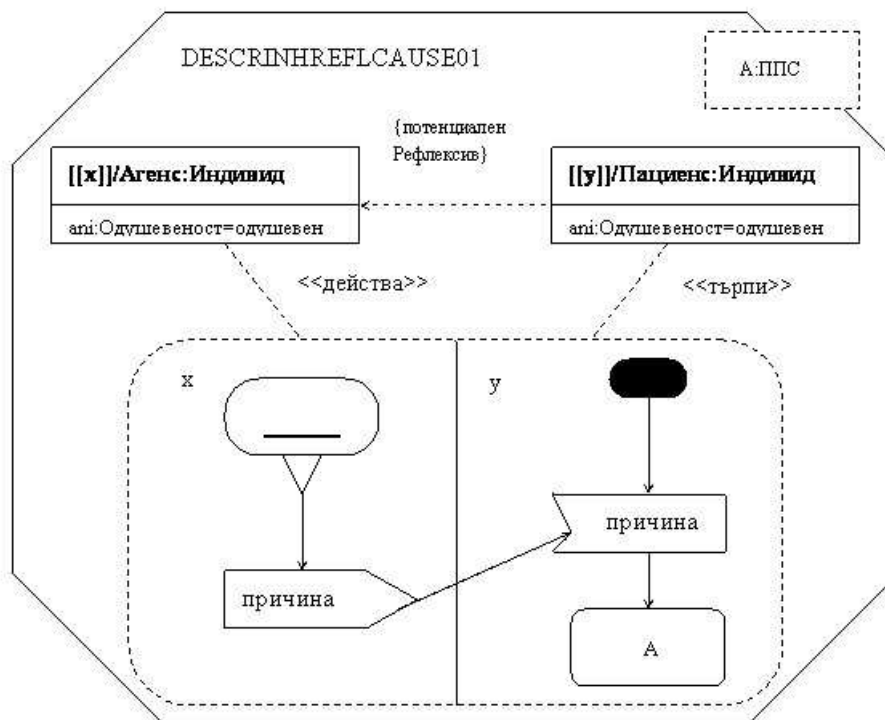
Each parent table is related to child tables, which contain the French and the Hungarian equivalents of the Bulgarian verbs. The relationship among the data is one-to-many, that is, to one Bulgarian verb there are one or more than one equivalents in French or Hungarian. In case of more than one equivalent in French or in Hungarian, the additional equivalents are given in separate tables. Figure 3 provides a schema of the database structure. The left-hand rectangle corresponds to the parent tables, and the right-hand rectangle designates the child tables. The first line in the rectangles indicates the table type, and the rest of the lines denote the fields of the respective table.

Descriptors of verb units with inherent reflexive meaning

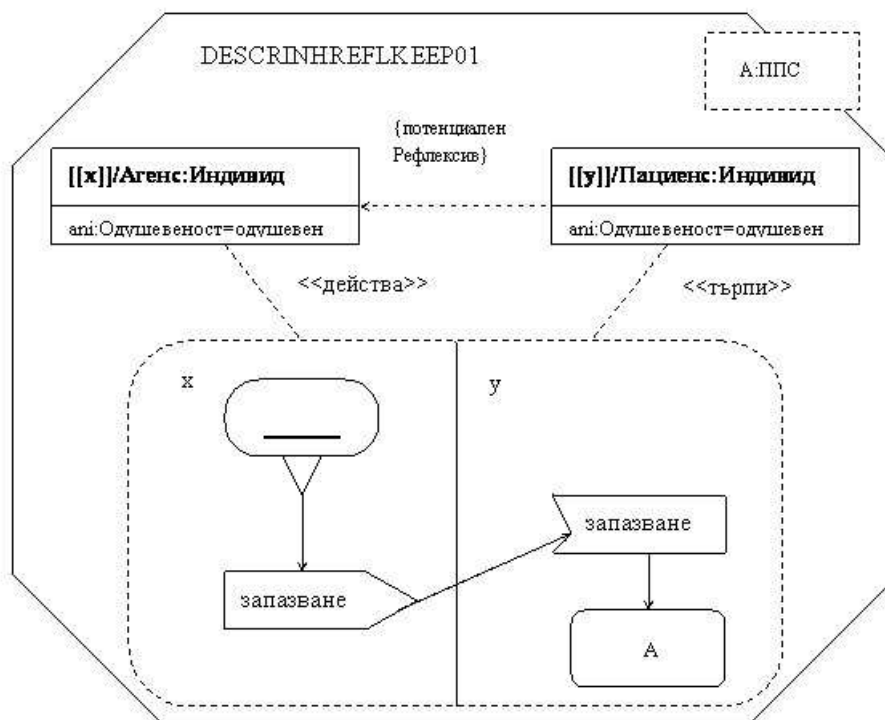
The inherent reflexives are modeled together with their non-reflexive correlatives using the **potentiallyReflexive** CONSTRAINT. Figures 4, 5 and 6 provide examples of EVENTIVITY FRAME TEMPLATES modeling the class of the inherent reflexive verbs. Examples of inherent reflexive verbs are: *задължавам* / *задължавам се* (*себе си*), фр. *obliger* / *s'obliger*, унг. *kötelez* / *kötelezi magát*, *улеснявам* / *улеснявам се* (*себе си*), фр. *aider* / *s'aider*, унг. *segít* / *segíti magát*, *виждам* / *виждам се* (*себе си*), фр. *voir* / *se voir*, унг. *lát* / *látja magát*; *виня* / *виня се* (*себе си*), фр. *accuser* / *s'accuser*, унг. *hibáztat* / *hibáztatja magát*.

Description of verb structures with reciprocal meaning

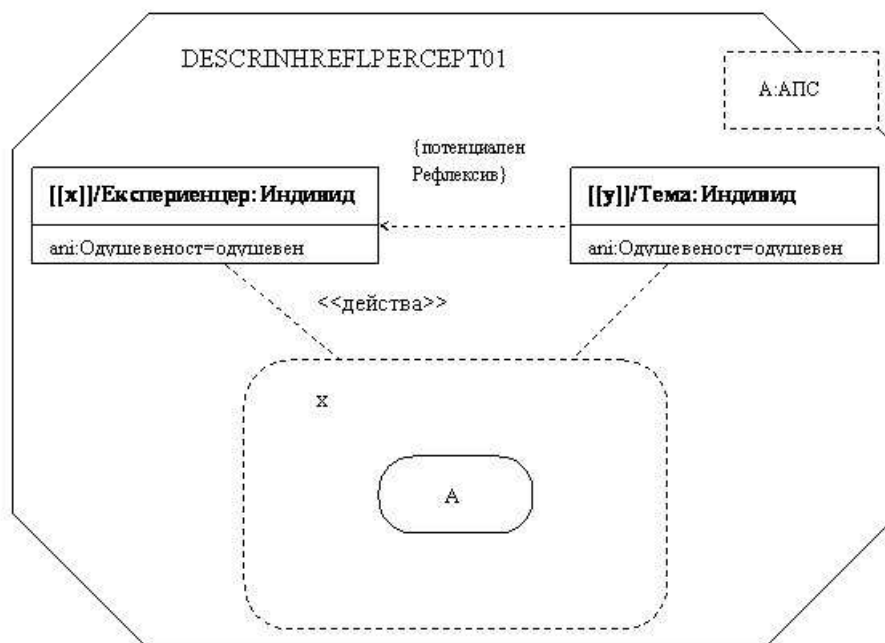
The eventivity frame templates of reciprocals are analogous to those of the inherent reflexives. The differentiating element is the **Reciprocal** CONSTRAINT, explained in Section 3. Examples of this class of verb structures are: *измъчваме се* (*един друг*) – фр. *se torturer* – унг. *kínózzák egymást*; *държим се* (*един друг*) – фр. *se tenir* – унг. *fogják egymást*;



Фиг. 4. Reflexive meaning with cause- SIGNAL- variant 1



Фиг. 5. Reflexive meaning with keep- SIGNAL- variant 1



Фиг. 6. Reflexive meaning - eventivity type *perception*

хвалям се (един друг) – фр. se vanter – унг. dicsérik egymást; виждаме се (един друг) – фр. se voir – унг. látják egymást; мразим се (един друг) – фр. se haïr – унг. gyűlölök egymást; обичаме се (един друг) – фр. s'aimer – унг. ismerik egymást; обвиняваме се (един друг) – фр. s'accuser – унг. vádolják egymást; познаваме се (един друг) – фр. se connaître – унг. ismerik egymást.

Descriptors of motive verbs

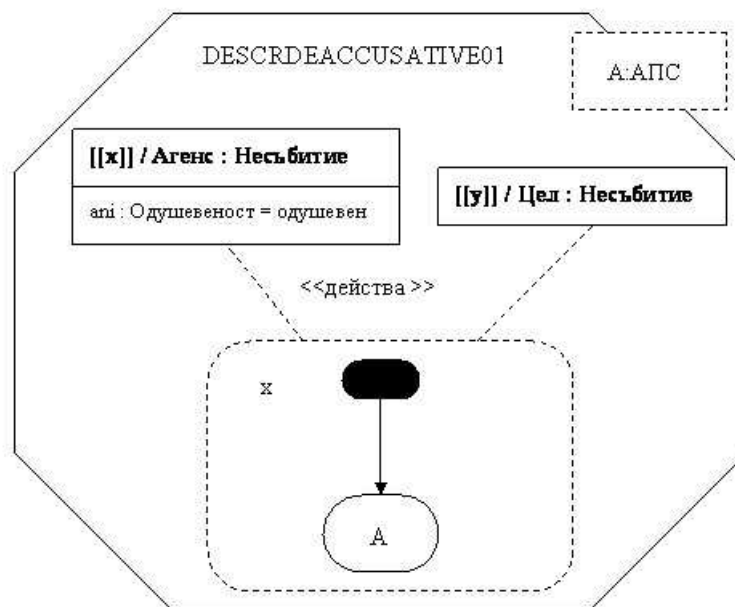
The verb units classified as motives are described by EVENTIVITY FRAMES with one participant in the **Agent** role, which moves or changes its location, for instance, *хвърлям се – фр. se jeter, se lancer – унг. veti magát; дървам се – фр. se tirer – унг. elhúzódik; издигам се – фр. s'élever – унг. felemelkedik.*

Descriptors of deaccusative verb structures

There are a small number of deaccusative verb structures in Bulgarian. Their equivalents in French are rarely deaccusative as well, for instance, *държа / държа се за – tenir / se tenir à, срещам / срещам се с – rencontrer / se rencontrer avec, отказвам / отказвам се от – refuser / se refuser à.* In Hungarian the deaccusative equivalents are even less in number, for instance, *опитвам / опитвам се – próbál / próbálkozik, държа / държа се за – fog / fogódzik.* The extent to which the meaning of the deaccusative verb deviates from that of its transitive counterpart differs for each verb. What is common is the focus on the **Agent** performing the action. Figure 7 is an example of an EVENTIVITY FRAME TEMPLATE describing deaccusative verbs.

Descriptors of absolutive verbs

The absolutive verbs are described by eventivity frames with one participant, which in



Фиг. 7. Descriptor of deaccusative verb structures - variant 1

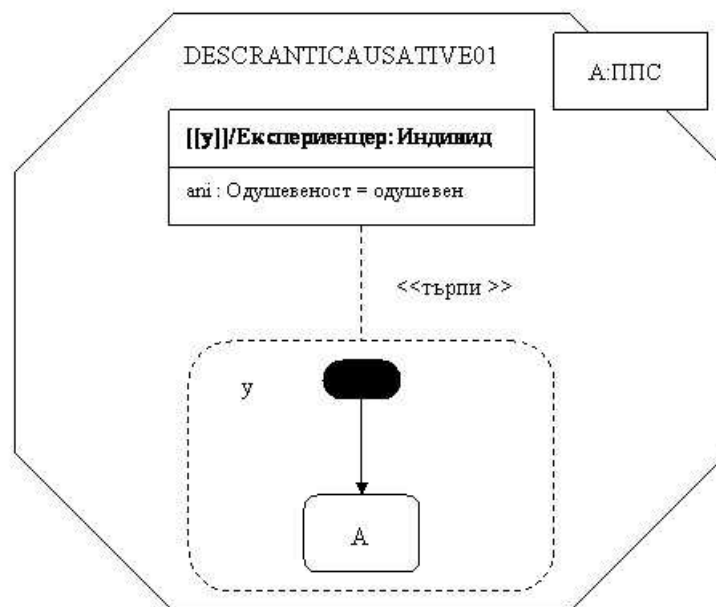
most of the cases is an **Agent**, for instance, *блъскам се* – фр. *se pousser, se bousculer* – унг. *lökdösődik, tolakodik*; *овладявам се* – фр. *se maîtriser, se contenir* – унг. *uralkodik magán*; *упражнявам се* – фр. *s'exercer* – унг. *gyakorolja magát*. In some cases the eventivity frame contains the abstract role CLASS *Instigator*, which can be instantiated either by the **Agent** or the **Effector** role, for instance, *вливам се* – фр. *se jeter* – унг. *ömlik, beömlik*; *подготвям се* – фр. *se préparer* – унг. *készül, felkészül*. The role of **Experiencer** is also possible, for instance, *чувствам се* – фр. *se sentir* – унг. *érzi magát*. In most of the cases the participant is animate.

Descriptors of anticausative verbs

In general, the action denoted by the anticausative verbs is perceived as "happening, triggered by itself". The focus is on the single prominent participant, which can be either animate or inanimate. Figure 8 provides an example of an **EVENTIVITY FRAME TEMPLATE** with the **Experiencer** PARTICIPANT ROLE, typical for the anticausatives. Examples of anticausative verbs are: *обиждам се* – фр. *s'offenser* – унг. *megsértődik*, *ядосвам се* – фр. *se fâcher, se mettre en colère* – унг. *bosszankodik*, *въдъхновявам се* – фр. *s'inspirer* – унг. *fellelkesedik*, *притеснявам се* – фр. *s'inquiéter* – унг. *nyugtalanodik, aggódik*, *събуждам се* – фр. *se réveiller, s'éveiller* – унг. *felébred*, *задушавам се* – фр. *s'étouffer* – унг. *fulladozik, megfullad*.

Descriptors of passive verb structures

The passive verb structures denote eventivities, in which the accent is on what happens with an inanimate prominent participant, for instance, *Книгите се продават на улицата* – фр. *Les livres se vendent dans la rue*; *Креслото се поставя пред камината* – фр. *Le fauteuil se place devant la cheminée*; *Подробност, която едва се забелязва* – фр. *Un détail qui s'aperçoit à peine*; *Понякога се наблюдава такава реакция* – фр. *Cette*



Фиг. 8. Descriptor of anticausative verbs - variant 1

réaction s'observe parfois. The abstract role CLASS *Instigated* with its elements **Patient** and **Theme** is typical for this type of entity.

Descriptors of *correlative(-)* verb units

It is difficult to find a common semantic feature for the verbs in this class, usually called *reflexiva tantum*. Synchronically, these verbs are considered arbitrary for each language, for instance, *боя се* - фр. *craindre*, *avoir peur* - унг. *fél*; *гордея се* - фр. *s'enorgueillir de*, *se glorifier de* - унг. *büszkélkedik*; *грижа се* - фр. *prendre soin de*, *soigner* - унг. *gondoskodik*; *усмихвам се* - фр. *sourire* - унг. *mosolyog*.

Contribution summary

The contribution of the thesis can be summarized as follows.

1. An overall infrastructure is proposed for representing the semantics of linguistic objects in a multilingual setting. An innovative object-oriented approach is applied, offering new possibilities for building optimized models and suitable language resources in the field of information technologies.
2. The morphosyntactic knowledge about verbs is systematized and represented according to the principles of formalization in computational linguistics.
3. A comparative investigation is carried out of significant for natural language processing types of verb units - reflexive verb structures in Bulgarian and their equivalents in French and in Hungarian.

4. The UER formalism is extended with new modeling elements necessary for representing the semantics of the selected language data.
5. *SemInVeSt*, a multilingual knowledge base of verbal semantics, is produced. *SemInVeSt* is a dynamic system, whose components are built on a service basis, that is, on a user demand and regarding the needs of a specific task.

Литература

- A. V. Bondarko, editor. *Functional Grammar Theory. Person. Voice*. Nauka Publishers, Sankt-Petersburg, 1991.
- A. V. Bondarko, editor. *Functional Grammar Issues: Field Structures*. Nauka Publishers, Sankt-Petersburg, 2005.
- Silvia Boteva. *The Verb in French and in Bulgarian. Functional-Semantic Grammar*. COLIBRI Publishing House, Sofia, 2000.
- Vyvyan Evans and Melanie Green. *Cognitive Linguistics: an Introduction*. Edinburgh University Press, UK, 2006.
- Vyvyan Evans, Benjamin Bergen, and Joerg Zinken. The cognitive linguistics enterprise: An overview. In *The Cognitive Linguistics Reader*, Advances in Cognitive Linguistics. Equinox, UK, 2008.
- Martin Fowler. *UML Distilled, Third Edition*. SoftPress Ltd., Sofia, 2004.
- Zlatka Guentchéva and Nicole Rivière. Reciprocal and reflexive constructions in French. In Vladimir P. Nedjalkov, editor, *Reciprocal Constructions*, volume 2 of *Typological Studies in Language 71*, chapter 12, pages 561–608. John Benjamins Publishing Company, 2007.
- George Lakoff. Cognitive models and prototype theory. In U. Neisser, editor, *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization*. Cambridge University Press, 1987.
- OMG. *OMG Unified Modeling Language Specification, Version 1.4*. Object Management Group (OMG), <http://www.omg.org/>, 2001.
- Elena Paskaleva. Compilation and validation of morphological resources (overview of the morphology cooking technologies). In Stelios Piperidis and Vangelis Karkaletsis, editors, *Proceedings of the International Workshop on Balkan Language Resources and Tools, First Balkan Conference on Informatics*, pages 68–74. Printed by INCOMA Ltd., Thessaloniki, Greece, 2003.
- Elena Paskaleva. *Computational Morphology. Resources and Tools*. Institute of Parallel Information Processing, BAS, Sofia, 2007.
- Elena Paskaleva, Kiril Simov, Mariana Damova, and Milena Slavcheva. The long journey from the core to the real size of large LDBs. In Peter Cole and Jerrold M. Sadock, editors, *Proceedings of a workshop sponsored by the special interest group on the lexicon of the Association for Computational Linguistics*, pages 161–169. Published by the Association for Computational Linguistics, Columbus, Ohio, USA, 1993.

- Jordan Penchev. Reciprocal and reflexive constructions in Bulgarian. In Vladimir P. Nedjalkov, editor, *Reciprocal Constructions*, volume 2 of *Typological Studies in Language* 71, chapter 13, pages 609–632. John Benjamins Publishing Company, 2007.
- Andrea C. Schalley. *Cognitive Modeling and Verbal Semantics. A Representational Framework Based on UML*. Trends in Linguistics. Studies and Monographs 154. Mouton de Gruyter, Berlin – New York, 2004.
- Kiril Simov, Petya Osenova, Milena Slavcheva, Sia Kolhovska, Elisaveta Balabanova, Dimitar Doikov, Krasimira Ivanova, Alexander Simov, and Milen Kouylekov. Building a linguistically interpreted corpus of Bulgarian: the BulTreeBank. In *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC 2002)*, pages 1729–1736, Canary Islands, Spain, 2002.
- Milena Slavcheva. Segmentation layers in the group of the predicate: a case study of Bulgarian within the BulTreeBank framework. In Erhard Hinrichs and Kiril Simov, editors, *Proceedings of the First Workshop on Treebanks and Linguistic Theories (TLT 2002)*, pages 199–210. Sozopol, Bulgaria, 2002.
- Milena Slavcheva. Some aspects of the morphological processing of Bulgarian. In *Proceedings of the Workshop on Morphological Processing of Slavic Languages, EACL 2003*, pages 71–77. Budapest, Hungary, 2003a.
- Milena Slavcheva. Corpus shallow parsing: Meeting point between paradigmatic knowledge encoding and syntagmatic pattern matching. In *Proceedings of the Corpus Linguistics Conference 2003*. Lancaster, UK, 2003b.
- Milena Slavcheva. Extracting verb complex structures in Bulgarian. In Hamish Cunningham, Elena Paskaleva, and Angelova Galia Bontcheva, Kalina, editors, *Proceedings of the International Workshop on Information Extraction for Slavonic and Other Central and Eastern European Languages, RANLP 2003, Borovets, Bulgaria*, pages 94–101. Context, Sofia, 2003c.
- Leonard Talmy. *Toward a Cognitive Semantics. Vol. 2: Typology and Process in Concept Structuring*. MIT Press, Cambridge, MA / London, 2000.