

Abstracts of Dissertations

Institute of Information and
Communication Technologies

BULGARIAN ACADEMY OF
SCIENCES



1 / 2012



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

Milena Slavcheva

СЕМАНТИЧНИ ДЕСКРИПТОРИ
НА РЕФЛЕКСИВНИ ПО ФОРМА
ГЛАГОЛНИ СТРУКТУРИ В
СЪВРЕМЕННИЯ БЪЛГАРСКИ,
ФРЕНСКИ И УНГАРСКИ ЕЗИК

Милена Славчева

Автореферати на дисертации

Институт по информационни и
комуникационни технологии

БЪЛГАРСКА АКАДЕМИЯ НА НАУКИТЕ

ISSN: 1314-6351

Поредицата „Автореферати на дисертации на Института по информационни и комуникационни технологии при Българската академия на науките“ представя в електронен формат автореферати на дисертации за получаване на научната степен „Доктор на науките“ или на образователната и научната степен „Доктор“, защитени в Института по информационни и комуникационни технологии при Българската академия на науките. Представените трудове отразяват нови научни и научно-приложни приноси в редица области на информационните и комуникационните технологии като Компютърни мрежи и архитектури, Паралелни алгоритми, Научни пресмятания, Лингвистично моделиране, Математически методи за обработка на сензорна информация, Информационни технологии в сигурността, Технологии за управление и обработка на знания, Грид-технологии и приложения, Оптимизация и вземане на решения, Обработка на сигнали и разпознаване на образи, Интелигентни системи, Информационни процеси и системи, Вградени интелигентни технологии, Йерархични системи, Комуникационни системи и услуги и др.

Редактори

Генадий Агре

Институт по информационни и комуникационни технологии, Българска академия на науките
И-мейл: agre@iinf.bas.bg

Райна Георгиева

Институт по информационни и комуникационни технологии, Българска академия на науките
И-мейл: rayna@parallel.bas.bg

Даниела Борисова

Институт по информационни и комуникационни технологии, Българска академия на науките
И-мейл: dborissova@iit.bas.bg

Настоящото издание е обект на авторско право. Всички права са запазени при превод, разпечатване, използване на илюстрации, цитирания, разпространение, възпроизвеждане на микрофилми или по други начини, както и съхранение в бази от данни на всички или част от материалите в настоящето издание. Копирането на изданието или на част от съдържанието му е разрешено само със съгласието на авторите и/или редакторите.

*The series **Abstracts of Dissertations of the Institute of Information and Communication Technologies at the Bulgarian Academy of Sciences** presents in an electronic format the abstracts of Doctor of Sciences and PhD dissertations defended in the Institute of Information and Communication Technologies at the Bulgarian Academy of Sciences. The studies provide new original results in such areas of Information and Communication Technologies as Computer Networks and Architectures, Parallel Algorithms, Scientific Computations, Linguistic Modelling, Mathematical Methods for Sensor Data Processing, Information Technologies for Security, Technologies for Knowledge management and processing, Grid Technologies and Applications, Optimization and Decision Making, Signal Processing and Pattern Recognition, Information Processing and Systems, Intelligent Systems, Embedded Intelligent Technologies, Hierarchical Systems, Communication Systems and Services, etc.*

Editors

Gennady Agre

Institute of Information and Communication Technologies, Bulgarian Academy of Sciences
E-mail: agre@iinf.bas.bg

Rayna Georgieva

Institute of Information and Communication Technologies, Bulgarian Academy of Sciences
E-mail: rayna@parallel.bas.bg

Daniela Borissova

Institute of Information and Communication Technologies, Bulgarian Academy of Sciences
E-mail: dborissova@iit.bas.bg

This work is subjected to copyright. All rights are reserved, whether the whole or part of the materials is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in other ways, and storage in data banks. Duplication of this work or part thereof is only permitted under the provisions of the authors and/or editor.



ABSTRACT OF PhD THESIS

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

Milena Nikolova Slavcheva

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 dissertation, 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 verb 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 encodes 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 the objects in reality (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 regards 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, characterizing the verb lexemes in a paradigmatic aspect (Slavcheva 2003a). The semantic descriptors in *SemInVeSt* are based on 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 verb semantics in a multilingual setting relates also to underlying principles of Functional Grammar and its functional-semantic categories. The work de-

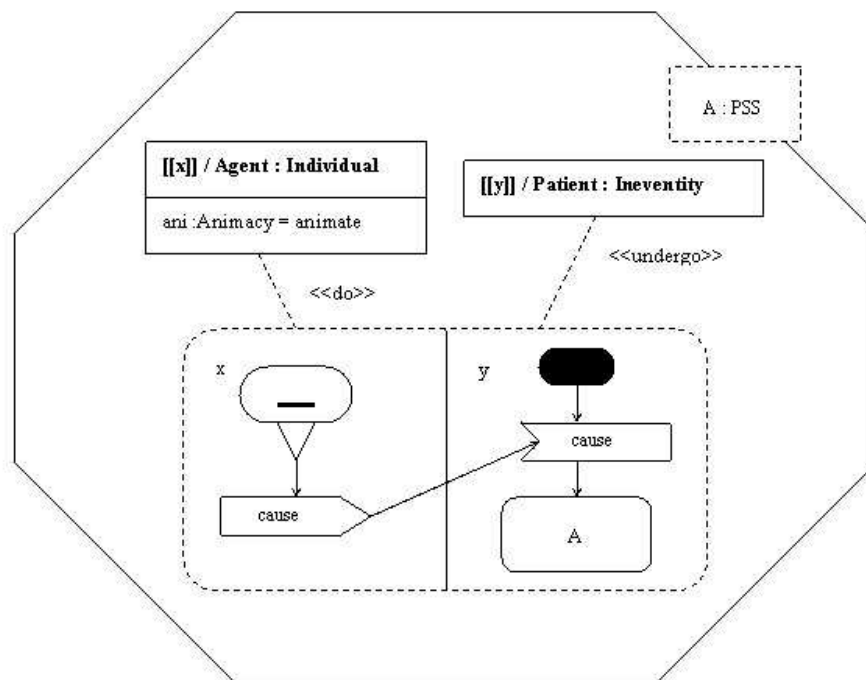


Fig. 1: TEMPLATE EVENTITY FRAME of a transitive verb

scribed here is particularly based on the theory of Functional Grammar developed by the Sankt-Petersburg linguistic circle. The theory has been applied in typological and contrastive studies of a multitude of languages (Bondarko 1991, 2005). The present classification of verbs is also compared to other contrastive and typological studies within the framework of Functional Grammar like (Boteva 2000; Penchev 2007; 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. Figure 1 (a generic diagram of the meaning of a transitive verb) illustrates 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 (Schalley 2004). 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 with rounded corners) is divided into SWIMLANES via a solid vertical line. Within the SWIMLANES, STATE-machines depict the participants' behaviour. In Figure 1 the left SWIMLANE belongs to the participant denoted by the *x* variable, and his 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, that is, it has a parameter to be bound.

2 Language objects described

In Bulgarian the language units described are character strings, which are a combination of a full-fledged verb form and a reflexive clitic *ce* or *cu*. Verb structures, containing the full-form of the reflexive personal pronoun *cebe cu*, are also described. *Cebe cu* is an unambiguous marker of prototypical, inherent reflexive meaning. Verb structures that contain the dative full-form pronoun *na cebe 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: *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), -kőd(ik); -kód(ik), -kőd(ik); -koz(ik), -kez(ik), -köz(ik); -kóz(ik), -kőz(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.), *въдъхвавам* (inspire s.o.) / *въдъхвавам 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), *страхувам 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 dissertation. 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, p.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 de-

note 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** (Schalley 2004). 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 eventivities 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 eventivities 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 EVENTIVITY FRAME.

Eventivity types and packages of modeling elements

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

Figure 2 provides the taxonomy of the *two-participant* eventivity type. In a *two-participant* eventivity, 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.

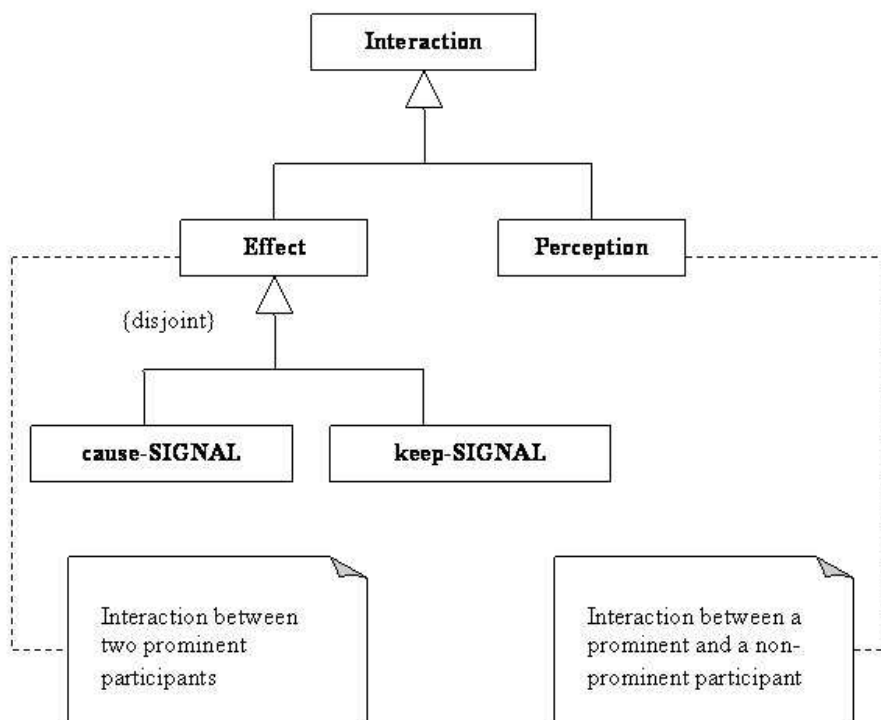


Figure 2: Entity types of two-participant *interaction*

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 his state. The impact, as well as the change of state, are taken in a general sense: physical impact and triggered state transition (bg. *изяждам*, fr. *manger*, hu. *megesz* (eat up); bg. *изсушавам*, fr. *sécher*, hu. *kiszárít* (dry up)), including movement, change of place (bg. *поставям*, fr. *placer*, hu. *helyez* (put, place); bg. *продвиждавам*, 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. *вдохновявам*, 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. *пробуцавам*, fr. *rallier*, *faire adhérer*, hu. *bevont* (join, involve)).

The second subtype of the *effect* eventivity 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 eventivities are conceptualized where 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. *őríz* (guard, protect, keep); bg. *съхранявам*, fr. *préserver*, *conserver*, hu. *megőríz*, *tartósít* (preserve, keep); bg. *задържам*, 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 – the *keep-SIGNAL*.

In the interaction between two participants, defined as *perception*, the eventivity 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. *пипам*, 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. *szerelem* (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 – the perceiver, and one non-prominent participant – the perceived.

In the *one-participant* eventivities, the participant is either in a given STATE or undergoes a STATE TRANSITION and there are variants of the participant's STATE-machine depending

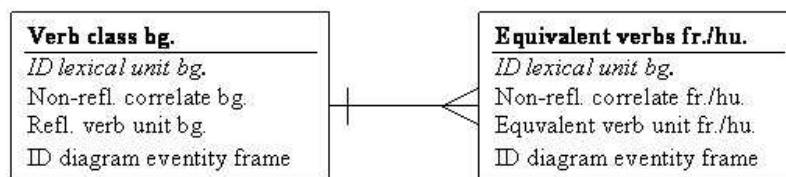


Figure 3: One-to-many relation between the data tables

on the presence or absence of a STATE TRANSITION, as well as on the types of source and target STATES.

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 EVENTITY 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 or 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 EVENTITY FRAME TEMPLATES modeling the class of the inherent reflexive verbs. Examples of inherent reflexive verbs are: bg. *зadължавам / задължавам се (себе си)*, fr. *obliger / s'obliger*, hu. *kötelez / kötelezi magát*; bg. *улеснявам / улеснявам се (себе си)*, fr. *aider / s'aider*, hu. *segít / segíti magát*; bg. *виждам / виждам се (себе си)*, fr. *voir / se voir*, hu. *lát / látja magát*; bg. *виня / виня се (себе си)*, fr. *accuser / s'accuser*, hu. *hibáztat / hibáztatja magát*.

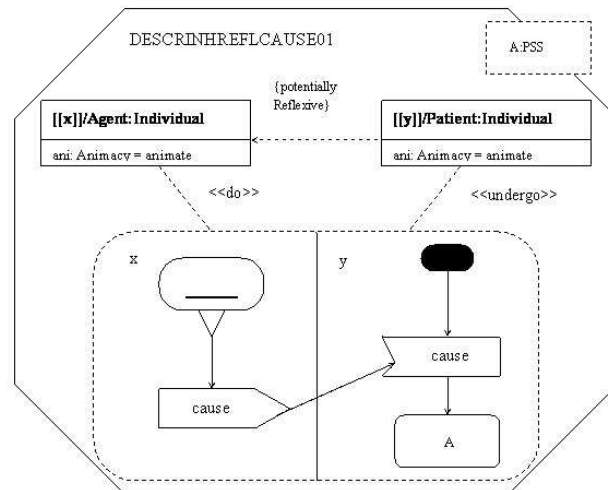


Figure 4: A reflexive meaning with cause-SIGNAL

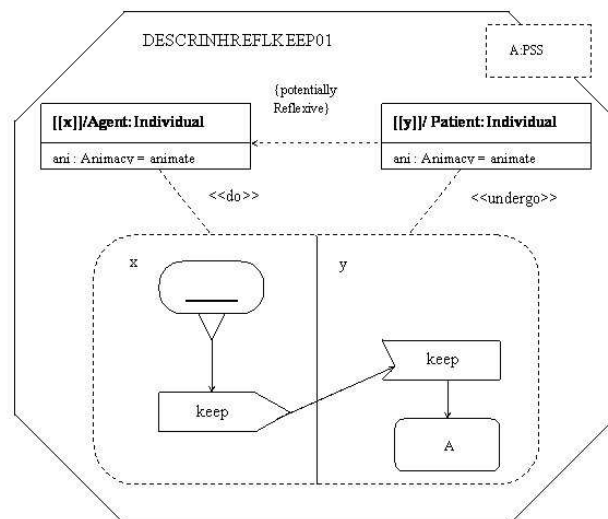
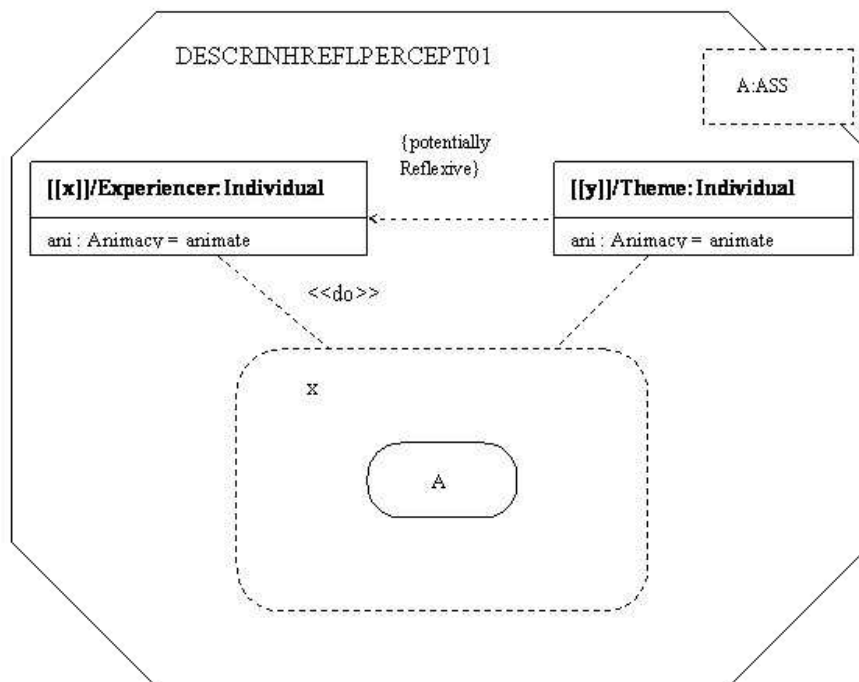


Figure 5: A reflexive meaning with keep-SIGNAL

Figure 6: A reflexive meaning of *perception*

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: bg. *измъчваме се (един друг)*, fr. *se torturer*, hu. *kínózzák egymást*; bg. *държим се (един друг)*, fr. *se tenir*, hu. *fogják egymást*; bg. *хвалим се (един друг)*, fr. *se vanter*, hu. *dicsérik egymást*; bg. *виждаме се (един друг)*, fr. *se voir*, hu. *látják egymást*; bg. *мразим се (един друг)*, fr. *se haïr*, hu. *gyűlölök egymást*; bg. *обучаваме се (един друг)*, fr. *s'aimer*, hu. *ismerik egymást*; bg. *обвиняваме се (един друг)*, fr. *s'accuser*, hu. *vádolják egymást*; bg. *познаваме се (един друг)*, fr. *se connaître*, hu. *ismerik egymást*.

Descriptors of motive verbs

The verb units classified as motives are described by EVENTIVITY FRAMES with one participant in the Agent role, who moves or changes his location, for instance, bg. *хвърлям се*, fr. *se jeter, se lancer*, hu. *veti magát*; bg. *дървам се*, fr. *se tirer*, hu. *elhúzódik*; bg. *издигам се*, fr. *s'élever*, hu. *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, bg. *държа / държа се за*, fr. *tenir / se tenir à*; bg. *срещам / срещам се с*, fr. *rencontrer / se rencontrer avec*; bg. *отказвам / отказвам се от*, fr. *refuser / se refuser à*. In Hungarian the deaccusative

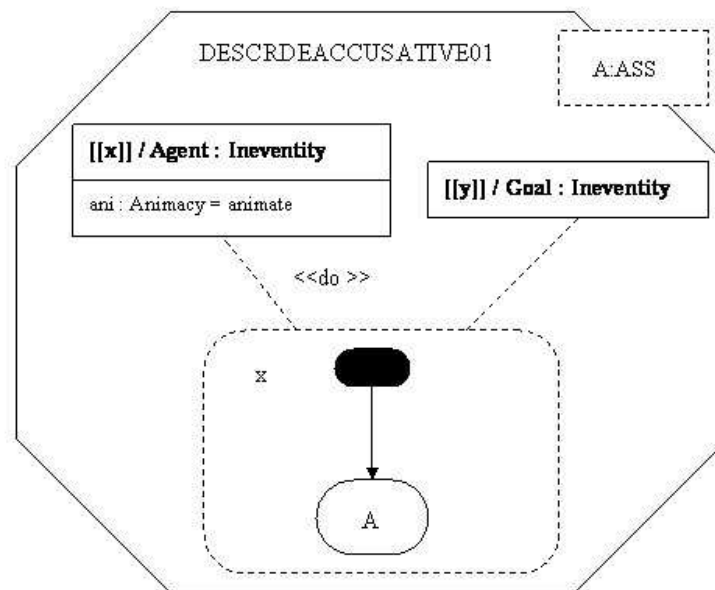


Figure 7: Descriptor of deaccusative verb structures

equivalents are even less in number, for instance, bg. *опитвам / опитвам се*, hu. *próbál / próbálkozik*; bg. *държа / държа се за*, hu. *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 **ENTITY FRAME TEMPLATE** describing deaccusative verbs.

Descriptors of absolute verbs

The absolute verbs are described by **ENTITY FRAMES** with one participant, who, in most of the cases, is an **Agent**, for instance, bg. *блъскам се*, fr. *se pousser, se bousculer*, hu. *lökdösődik, tolakodik*; bg. *овладявам се*, fr. *se maîtriser, se contenir*, hu. *uralkodik magán*; bg. *упражнявам се*, fr. *s'exercer*, hu. *gyakorolja magát*. In some cases the **ENTITY FRAME** contains the abstract role **CLASS Instigator**, which can be instantiated either by the **Agent** or the **Effector** role, for instance, bg. *вливам се*, fr. *se jeter*, hu. *ömlik, beömlik*; bg. *подготвям се*, fr. *se préparer*, hu. *készül, felkészül*. The role of **Experiencer** is also possible, for instance, bg. *чувствам се*, fr. *se sentir*, hu. *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, who can be either animate or inanimate. Figure 8 provides an example of an **ENTITY FRAME TEMPLATE** with the **Experiencer PARTICIPANT ROLE**, typical for the anticausatives. Examples of anticausative verbs are: bg. *обиждам се*, fr. *s'offenser*, hu. *megsértődik*; bg. *ядосвам се*, fr. *se fâcher, se mettre en colère*, hu. *boszankodik*; bg. *вдъхновявам се*, fr. *s'inspirer*, hu. *fellelkesedik*; bg. *притеснявам се*, fr. *s'inquiéter*, hu. *nyugtalanodik*,

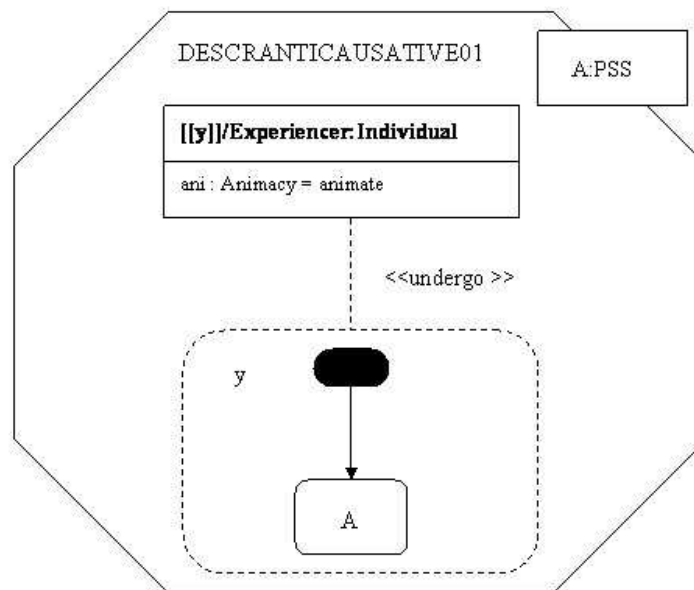


Figure 8: Descriptor of anticausative verbs

aggódik; bg. *събуждам се*, fr. *se réveiller, s'éveiller*, hu. *felébred*; bg. *задушавам се*, fr. *s'étouffer*, hu. *fulladozik, megfullad*.

Descriptors of passive verb structures

The passive verb structures denote eventities, in which the accent is on what happens with an inanimate prominent participant, for instance, bg. *Книгите се продават на улицата*, fr. *Les livres se vendent dans la rue*; bg. *Креслото се поставя пред камината*, fr. *Le fauteuil se place devant la cheminée*; bg. *Подробност, която едва се забелязва*, fr. *Un détail qui s'aperçoit à peine*; bg. *Понякога се наблюдава такава реакция*, fr. *Cette réaction s'observe parfois*. The abstract role CLASS *Instigated* with its elements *Patient* and *Theme* is typical for this type of eventity.

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, bg. *боя се*, fr. *craindre, avoir peur*, hu. *fél*; bg. *гордея се*, fr. *s'enorgueillir de, se glorifier de*, hu. *büszkélkedik*; bg. *грижа се*, fr. *prendre soin de, soigner*, hu. *gondoskodik*; bg. *усмихвам се*, fr. *sourire*, hu. *mosolyog*.

Contribution summary

The contribution of the dissertation 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 appropriately formalized for use in the field of 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 verb 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.

References

- 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.



АВТОРЕФЕРАТ НА ДИСЕРТАЦИЯ

за присъждане на образователна и научна степен „доктор“
по научна специалност 05.04.11 „Общо и сравнително
езикознание (математическа лингвистика)“

СЕМАНТИЧНИ ДЕСКРИПТОРИ НА РЕФЛЕКСИВНИ ПО ФОРМА ГЛАГОЛНИ СТРУКТУРИ В СЪВРЕМЕННИЯ БЪЛГАРСКИ, ФРЕНСКИ И УНГАРСКИ ЕЗИК

Милена Николова Славчева

Ръководител: доц. д-р Георги Гаргов

Научно жури:

проф. дн Иван Касабов
доц. д-р Мария Стамболиева
доц. д-р Елена Паскалева
доц. дн Галя Ангелова
доц. д-р Кирил Симов



Дисертацията е обсъдена и допусната до защита на разширено заседание на секция „Лингвистично моделиране“ на ИИКТ-БАН, състояло се на 10 май 2011г.

Защитата на дисертацията е проведена на 25 октомври 2011 г. от 11:00 часа в зала 218 на блок 25А, ИИКТ-БАН.

Пълният обем на дисертацията е 210 страници. Състои се от увод и пет глави (стр. 1-161), включва и приложение (стр. 176-210). Списъкът на използваната литература съдържа 145 заглавия (стр. 162-175). В текста на дисертацията са включени 37 таблици и 84 фигури, чиито списъци са дадени в началото на дисертацията.

Увод

Обект на изследване в дисертационния труд е семантиката на *se/ci* - глаголи в българския език и техни семантични еквиваленти във френския и унгарския език. Изследването е в областта на компютърната лингвистика и езиковите технологии и е свързано с прилагането на обектно-ориентиран подход при представяне на езиковите единици.

Целта на разработката е създаването на формализирани семантични описания на глаголните единици като компонент от цялостна инфраструктура за представяне на глаголната семантика в многоезиков план.

За постигане на така формулираната цел е необходимо да бъдат решени следните конкретни задачи:

1. Дефиниране на класове глаголи в българския език, които са изходните единици в съпоставката с глаголи във френския и унгарския език.
2. Дефиниране на пакети от моделиращи елементи, необходими за описание на избраните езиковите обекти на концептуално равнище.
3. Разширяване на изходния модел с нови моделиращи елементи, необходими за семантичното описание на езиковите данни от настоящата разработка.
4. Систематизиране на езиковите данни и построяване на диаграмите - семантични дескриптори.

Стремежът е да се построят формални модели, които са максимално лингвистично мотивирани и аргументирани. Технологичната реалност, обаче, налага да се направи разумен компромис със строгата теория за езика. Трябва да се подчертае, че компромисите са необходими най-вече при езикови категории и явления, които са проблемни и остават неразрешени и в самата лингвистична теория. В настоящата разработка е залегнало разбирането, че **която и да е класификация е изкуствено построение въз основа на преливащи се едно в друго лингвистични явления**. Такива построения са необходими за създаването на езикови модели и езикови ресурси в компютърната лингвистика.

Основен проблем, с който се сблъскваме в лингвистичните изследвания, е употребата на езика като средство за описание на самия език, или **метаезика** на езиковедските изследвания. Тук в редица случаи е неизбежна **условността в употребата на лингвистичните термини**. В заглавието на дисертационния труд е употребено условно обобщаващо наименование *рефлексивни по форма глаголни структури* за да се обозначат по сравнително удобен начин твърде разнородните по своята семантика глаголни единици. Във формално-структурно отношение глаголните единици също са разнородни в разглежданите езици и представляват символни низове, включващи специфични елементи (клитики, морфеми, самостоятелни думи), които в едни случаи се проявяват отчетливо като рефлексивни елементи, а в други случаи се отдалечават от рефлексивното значение. Ясно е, че само част от такива глаголни единици са носители на същинско рефлексивно значение – вършителят на някакво действие въздейства не върху друг обект, а върху самия себе си, или, в по-широк смисъл, из-

вършва действие за себе си. Голяма част от така наречените рефлексивни по форма глаголни структури изразяват широк спектър от нюансирани значения. Тези значения образуват континуум между полюса на прототипното рефлексивно значение и противоположния му полюс на отсъстващо рефлексивно значение. Необходимо е да се подчертае, също така, че терминът *структура* тук се употребява в неговото най-общо значение – той назовава значими за настоящото многоезиково изследване глаголни единици (разнородни по състав в различните езици), а не глаголни фрази, които са обект на изследване в синтактични формални теории. Тъй като фокусът на дисертационния труд е върху концептуалните структури, които намират израз в различни езикови форми, в изложението е въведен и терминът *композиция* за обозначаване на по-широките от една ортографична дума езикови цялости, с цел да се избегне аналогията с формално дефинираните глаголни фрази (Verb Phrases (VPs)) в синтактичните теории.

Настоящото изследване е когнитивно ориентирано – основен обект на анализ е семантичната концептуализация, което означава, че **семантичните параметри се приемат за доминиращи над морфосинтактичните**. Тук възниква теоретичният (дори философският) въпрос за това до каква степен е възможно да се абстрахираме от граматичната структура когато се стремим да построим когнитивно мотивирани семантични дескриптори на езикови единици. Или въпросът е доколко езикът като относително автономна система може да оказва обратно въздействие върху когнитивните процеси. Настоящата разработка се основава на виждането, че между мисленето, езика и света има тясно взаимодействие и многопосочно въздействие един върху друг.

На езиковите единици се съпоставят семантични модели, съобразени с теорията и формализма на **Единно представяне на събития (ЕПС)** (Unified Eventy Representation (UER)) (Schalley 2004). Подходът на ЕПС към глаголната семантика е когнитивен – описват се **събития** (eventities), които, най-общо казано, представляват протичащи във времето явления. ЕПС е **графичен формализъм**, който чрез **диаграми** представя концептуалната структура на събитията, която е съвкупност от 1) статични моделиращи елементи, характеризиращи участниците в събитието и асоциативните релации между тях; 2) динамични моделиращи елементи, описващи поведението на участниците и взаимодействието между тях чрез механизми на състоянията (state-machines). В диаграмите на ЕПС се използва комбинация от графични елементи и линейни текстови конструкции. Апаратът за моделиране на лингвистичното знание в ЕПС е приложение, адаптиране и разширение на **Unified Modeling Language (UML)** (Единен език за моделиране) – международен стандарт за графично представяне, моделиране и дизайн на **обектно-ориентирани** системи в областта на информатиката (OMG 2001), (Фаулър 2004).

Продуктът от разработката е база от знания за семантиката на лексикални езикови обекти – глаголи в българския език, на които са дадени семантични еквиваленти във френския и в унгарския език. Базата от знания за семантиката на глаголите се нарича **Семинвест**, което е съкращение от пълното наименование на езиковия ресурс – **Семантично интерпретирани вербални (глаголни) структури**. **Семинвест** се отличава съществено от традиционните речници по характера на съдържащите

се в него езикови единици. Те представляват разнородни по състава си конструкции от морфосинтактична гледна точка, които, обаче, се асоциират със семантични цялости от концептуална гледна точка. Освен това **Семинвест** е динамична система от структури – той се изгражда според предпочитания напоследък в областта на информационните технологии принцип на услуги по заявка на потребител, т.е. изграждане и предоставяне на компоненти съобразно конкретна задача, а не предоставяне на готов продукт, проектиран с определена изчерпателност на езиковите данни и лингвистично знание за тях.

Формализираното представяне на знания за семантиката на езиковите обекти има дългогодишна традиция в България. Значителни са, например, постиженията в тази област на *Секцията за лингвистично моделиране* (СЛМ), ИККТ - БАН, където в редица проекти под ръководството на Галя Ангелова и Кирил Симов се разработват семантични модели, семантични езикови ресурси и компютърни системи, базирани на семантични знания (вж. напр. (Angelova 2000; Boytcheva and Angelova 2001; Angelova 2005), (Simov 2006; Simov and Osenova 2008), както и информацията и богатата библиография на сайта на СЛМ¹ и на VulTreeBank²). В индустриалния сектор може би най-успешната източноевропейска фирма в областта на информационните технологии OntoText (под ръководството на Атанас Киряков) провежда изследвания и разработва технологии със семантична насоченост (вж. напр. (Popov et al. 2003; Kiryakov 2010, 2006), както и дългия списък на публикации на сайта на Ontotext³).

1. Цели, методи и изследователски подходи

В настоящата разработка се прилага теория и формализъм за описание на глаголната семантика, чийто подход е когнитивен – значението на глаголите се разглежда като неразривно свързано с концепти за събития и подобни мисловни единици. Когнитивният аспект на лингвистичния анализ се изразява в наличието на общи ръководни принципи, отнасящи се до връзката между възприятията, мисленето и езика – между концептуалните структури, значенията и формите на езиковите единици. Общите теоретични принципи се основават на схващането, че езикът отразява мисловни структури (patterns of thought), затова изучаването на езика е свързано с изучаването на концептуални структури (patterns of conceptualization) (Evans and Green 2006). Езикът кодира (code) и предава (transmit) мисли чрез символи (symbols) – езикови единици, които могат да бъдат части от думи (морфеми), думи, или комбинации от думи (изрази). Символите се състоят от форма и асоциирано с нея значение (form-meaning pairing), или те са символни комбинации (symbolic assemblies). Значението, като съставна част на символната комбинация, е свързано с мисловни образи, представи, които се наричат понятия, или концепти (concepts). Концептите се пораждат от възприятията, или от перцептите. Перцептите са резултат от взаимодействието на човека със заобикалящия го свят, те представляват разбирането на човека за свойст-

¹<http://lml.bas.bg/>

²<http://www.bultreebank.org/>

³<http://www.ontotext.com/publications>

вата на обектите от действителността (Lakoff 1987). Перцептите сформират ментални образи (mental images) – претворената от човешкото съзнание реалност.

Настоящата разработка върху глаголната семантика има своите опорни точки в морфосинтактичното описание на глаголните лексеми. Един от източниците на структурирано морфосинтактично знание за така наречените рефлексивни по форма глаголни единици в българския език произлиза от участието ми в работата по създаването на голяма лексикална база на българския език (Paskaleva et al. 1993), (Паскалева 2007) и по-специално върху признаците, характеризиращи глаголните лексеми в парадигматичен план (Slavcheva 2003a), (Славчева 2003). Изграждането на семантичните дескриптори в настоящата разработка се основава на семантичния анализ на глаголни лексеми от така наречения *Кратък компютърен граматичен речник*⁴, създаден в *Секцията за лингвистично моделиране (СЛМ), Институт по паралелна обработка на информация (ИПОИ)* към БАН (Paskaleva 2003). Речникът съдържа 10 216 лексеми и е съставен въз основа на честотни списъци, извлечени чрез морфологичен анализ на текстови корпуси (с обем 5 милиона думи). Лексикалният състав на **Семинавест** за българския език е извлечен от 2465 глагола от *Краткия компютърен граматичен речник* и допълнен с глаголни лексеми от голямата база. Понастоящем са представени 330 глаголни единици на български език и техните еквиваленти на френски и унгарски език.

Друг източник на лингвистично знание за *се/си* - глаголите в българския език е свързан с работата ми по създаване и прилагане на граматика за частичен синтактичен автоматичен анализ, които маркират в текста сегменти, представляващи така наречения *глаголен комплекс* в синтагматичен план – низ от спомагателни глаголи, местоименни клитики, частици и пълнозначен глагол (Simov et al. 2002; Slavcheva 2002, 2003b,c).

В настоящата разработка моделирането на глаголната семантика в многоезиков план се опира на основни принципи на *функционалната граматика* и на нейни *функционално - семантични категории* за да се дефинират релевантни класове глаголни единици в трите езика. Настоящата разработка се основава на разработваната от Петербургската школа теория на функционалната граматика, приложена в типологични и съпоставителни изследвания на множество езици (Бондарко 1991, 2005). Настоящата класификация на глаголните единици се основава също така на съпоставката на глаголите във френския и българския език според принципите на функционалната граматика, представена от С. Ботева (Ботева 2000), както и на типологичните изследвания на глаголни конструкции в българския и френския език от гледна точка на функционалната граматика, представени съответно в (Penchev 2007) и (Guentchéva and Rivière 2007).

Изборът на френски и унгарски като езици за съпоставка с български е обусловен от следните фактори. Във френския и в българския език има сходство в проявенията на изследваните глаголни единици, но има и разлики. А съпоставянето на езици като френския и българския, които имат общ индоевропейски произход, но са и достатъчно различни, наистина позволява да се направят много интересни и ценни

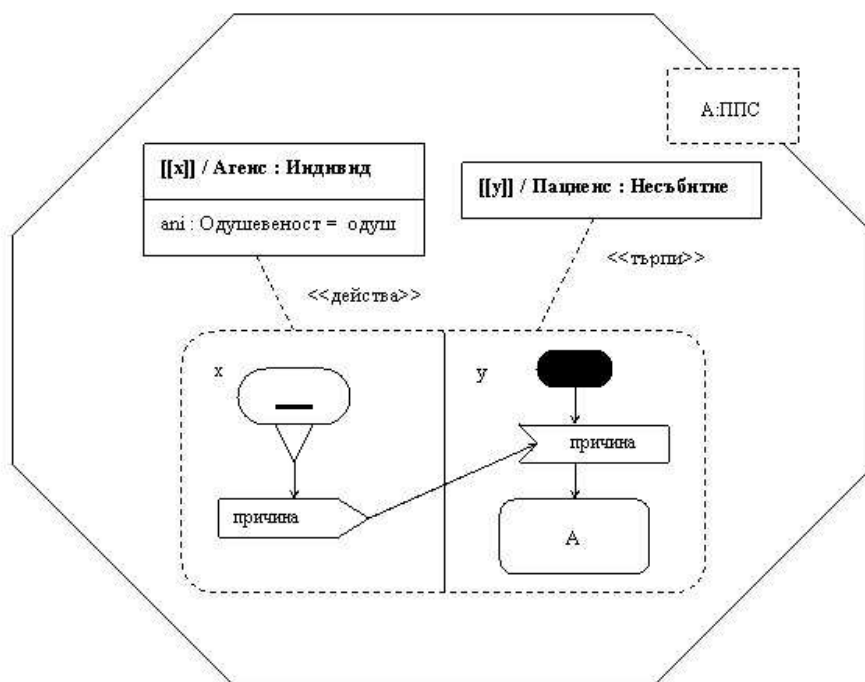
⁴Речникът е достъпен в Интернет: <http://www.larflast.bas.bg/balric/index/index.htm>

наблюдения и изводи, до които не би довело проучването само на един език. Що се отнася до избора на унгарския език, при сравняването на езици с малко допирни точки, в случая един угро-фински и един индоевропейски език, на преден план изпъкват глобалните, очевидните разлики, но така пъзелът на езиковото явление се допълва с ценни елементи.

В настоящата разработка прилагам теорията и формализма *Единно представяне на събития* (ЕПС) (Schalley 2004) с цел построяване на семантичните дескриптори на глаголните единици. Апаратът за моделиране на лингвистичното знание в ЕПС е приложение, адаптиране и разширение на *Unified Modeling Language (UML)* (Единен език за моделиране). ЕПС е графичен формализъм, който чрез диаграми представя концептуалната структура на събитията. В диаграмите се използва комбинация от графични елементи и линейни текстови конструкции. Диаграмите в ЕПС са съставени от свързани чрез пътеки (paths) възли (nodes). Основно се използват четири вида графични конструкти: 1) икони (фигури с фиксиран размер, които не могат да включват съдържателни изрази (линейни моделиращи елементи)); 2) двумерни символи (фигури с варираща височина и ширина, които могат да съдържат други моделиращи елементи); 3) символни низове (представят различни видове информация в езикова форма или логически запис); 4) пътеки (последователности от линии, чиито краища са свързани или едно към друго, или към двумерен символ, или към икона). На Фиг. 1 е представена една обобщена диаграма на значението на преходен глагол за да бъде илюстриран най-общо апаратът за описание на глаголната семантика.

Основният концепт събитие е представен чрез СЪБИТИЙНА РАМКА (EVENTITY FRAME), чиято графична нотация е осмоъгълник, съдържащ останалите графични елементи, обозначаващи концептите - компоненти на събитието. СЪБИТИЙНАТА РАМКА е съставена от динамично ядро (dynamic core) и статична периферия (static periphery). Динамичното ядро описва графично динамичния аспект от концептуалната структура на събитието и се представя чрез очертан с пунктир правоъгълник със заоблени краища. Статичният аспект на събитието е представен чрез графичните елементи, разположени около динамичното ядро в СЪБИТИЙНАТА РАМКА. Основни елементи от статичната периферия са участниците в събитието, които в диаграмата са представени от КЛАСОВЕ УЧАСТНИЦИ (PARTICIPANT CLASSES) – на Фиг. 1 това са двата правоъгълника, разположени в горната част на ограждащия осмоъгълник.

Динамичното ядро на е разделено на две части чрез плътна вертикална линия. Всяка такава част е отделението, КОРИДОРЪТ (SWIMLANE), в който се съдържа описанието на поведението (behaviour) на даден участник. Левият КОРИДОР на Фиг. 1 е моделът на поведението на участника x, а десният – на участника y. Моделиращите елементи, които пресичат плътната вертикална линия (границата на КОРИДОРИТЕ) представят взаимодействието (interaction) между участниците. Участник x извършва действие (action), представено чрез АКТИВНО ПРОСТО СЪСТОЯНИЕ (ACTIVE SIMPLE STATE), чиято нотация е фигура с прави горна и долна линия и изпъкнали дъги от двете страни. СЪСТОЯНИЯТА могат да имат имена, но в случая долната черта във вътрешността на АКТИВНОТО ПРОСТО СЪСТОЯНИЕ означава, че то не е специфицирано, т.е. не се знае точно какво действие извършва участникът x. Знае се, обаче,



Фиг. 1. Транзитивна СЪБИТИЙНА РАМКА - ШАБЛОН

че това действие причинява ПРЕХОД на СЪСТОЯНИЕТО (state transition) на участник y , изобразен в десния КОРИДОР. Причиняването на промяната в състоянието на участник y се моделира от СИГНАЛ-причина (cause-SIGNAL), чиято нотация се състои от два петогълника и свързваща стрелка, която пресича вертикалната линия - разделител на КОРИДОРИТЕ на участниците. От петогълника в КОРИДОРА на участник x се изпраща СИГНАЛ, който се приема от петогълника в КОРИДОРА на участник y . Промяната в състоянието на участник y се представя от ПРЕХОД, изобразен в десния КОРИДОР на динамичното ядро. ПРЕХОДЪТ се състои от изходно и крайно СЪСТОЯНИЕ, свързани със стрелка, която в разглеждания пример е разделена поради получаването на СИГНАЛ-причина, пораждащ този ПРЕХОД. Крайното СЪСТОЯНИЕ е ПАСИВНО ПРОСТО СЪСТОЯНИЕ, чиято нотация е правоъгълник със заоблени краища. В разглеждания пример крайното състояние е параметризиран елемент с параметър A , което означава, че СЪБИТИЙНАТА РАМКА на Фиг. 1 е ШАБЛОН (TEMPLATE). ШАБЛОНЪТ е параметризиран моделиращ елемент, т.е. съдържа един или повече параметри, на които трябва да бъдат дадени стойности. Нотацията на ШАБЛОНА е очертан с пунктир малък правоъгълник, който съдържа списък на параметрите и техните типове. Правоъгълникът се разполага в горния десен ъгъл на моделиращия елемент, който съдържа параметри, в разглеждания пример това е СЪБИТИЙНАТА РАМКА. В правоъгълника е изписан изразът $A:ППС$ – посочен е параметърът A и неговият тип ППС (пасивно просто състояние). Изходното СЪСТОЯНИЕ на ПРЕХОДА в десния КОРИДОР (плътна фигура) не е специфицирано

нито съдържателно (чрез име на СЪСТОЯНИЕТО, например, Чист, Празен и т.н.), нито структурно (чрез формата на фигурата, означаваща типа на СЪСТОЯНИЕТО като активно, пасивно и т.н.).

Участниците в събитието са представени в статичната периферия на СЪБИТИЙНАТА РАМКА чрез КЛАСОВЕ УЧАСТНИЦИ (PARTICIPANT CLASSES), изразени чрез линейни моделиращи елементи, оградени в правоъгълници. Синтаксисът на линейния моделиращ елемент е следният: *[[представител на участници]] / роля на участник : тип на участник*. В примера на Фиг. 1 това са изразите *[[x]] / Агенс : Индивид* и *[[y]] / Пациенс : Несъбитие*, оградени с правоъгълен контур. ПРЕДСТАВИТЕЛЯТ НА УЧАСТНИЦИТЕ е променлива (x и y на Фиг. 1), която обозначава потенциални участници, които отговарят на ограниченията, посочени в КЛАСА УЧАСТНИЦИ. РОЛЯТА НА УЧАСТНИК е израз, който дефинира семантична роля, например, Агенс, Пациенс. ТИПЪТ УЧАСТНИК посочва онтологичната категория на участника, например, Несъбитие, Индивид. Долното отделение на правоъгълника, обозначаващ даден участник, е мястото на АТРИБУТИТЕ, характеризиращи допълнително участника. На Фиг. 1 АТРИБУТЪТ, характеризиращ участник x има име *ani* и е от тип (data type) *Одушевеност* със стойност *одушевен*.

КЛАСОВЕТЕ УЧАСТНИЦИ са свързани с динамичното ядро чрез РЕЛАЦИИ НА УЧАСТИЕ (PARTICIPATE ASSOCIATIONS), изразени графично чрез пунктирани линии между правоъгълниците на КЛАСОВЕТЕ УЧАСТНИЦИ и динамичното ядро. В нашия пример РЕЛАЦИИТЕ НА УЧАСТИЕ са специфицирани чрез СТЕРЕОТИПИТЕ «действия» («do») и «търпи» («undergo»), които маркират макроролите (или проторолите) на главните участници - съответно *действащ* (doer) и *търпящ* (undergoer).

2. Езикови обекти на изследване

Обект на изследване в българския език са символни низове, които са комбинация от глаголна форма и клитика *се* или *си*. Описват се и глаголни единици с пълната форма на възвратното лично местоимение *себе си*, което е еднозначен изразител на прототипно, или същинско рефлексивно значение. Разглеждат се и глаголни единици, включващи дативното местоимение *на себе си*.

Глаголните единици във френския език, които са обект на изследване, представляват комбинация от глаголна форма и несамостоятелно лично местоимение (pronот соjoint), което само във формата *se* за трето лице единствено и множествено число е възвратно. Другите му форми съвпадат с тези на личните местоимения и се изменят по лице и число: *me, te, nous, vous: Je me promène* (Аз се разхождам), *Tu te promènes* (Ти се разхождаш), *Elle se promène* (Тя се разхожда), *Nous nous promenons* (Ние се разхождаме), *Vous vous promenez* (Вие се разхождате), *Elles se promènent* (Те се разхождат).

Обект на изследване в унгарския език са глаголи, които съдържат рефлексивен по характер суфикс от следното множество: *-ód(ik), -őd(ik); -kod(ik), -ked(ik), -kőd(ik); -kőd(ik), -kőd(ik); -koz(ik), -kez(ik), -köz(ik); -kóz(ik), -köz(ik); -óz(ik), -öz(ik); -ódz(ik), -ődz(ik); -oz(ik), -ez(ik), -öz(ik); -edz(ik); -ul, -ül*. Няколко други наставки също могат

да придадат рефлексивна отсянка на образуваните с тях глаголи: девербалните *-ad*, *-ed* и деноминалните *-z*, *-l*, *-ll*. Рефлексивни глаголни композиции в унгарския език се образуват и с възвратното местоимение *magá*, което се мени по лице и число: *magam* - 1 л., ед. ч., *magad* - 2 л., ед. ч., *magá* - 3 л., ед. ч., *magunk* - 1 л., мн. ч., *magatok* - 2 л., мн. ч., *maguk* - 3 л., мн. ч.

Езиковите обекти, които се изследват и описват в настоящата разработка, са особено трудно поддаващи се на класифициране. В дисертацията е предложена една класификация, която разпределя езиковите единици от трите съпоставяни езика в условно зададени групи, които служат за ориентири при изграждането на семантичните дескриптори на глаголните единици. Тази класификация не е абсолютна, тя не е и крайна цел на разработката.

Темата за класовете на рефлексивните глаголни единици придобива значение първо поради наличието на опозицията рефлексивни и неререфлексивни форми на глаголите. Критерий, спрямо който се подразделят така наречените рефлексивни по форма глаголни единици, е *съотносителността* с неререфлексивни по форма глаголни единици. Съществуването или несъществуването на съотносителна неререфлексивна единица определя по-нататъшния анализ и класификация на рефлексивните по форма глаголни единици. Стойностите на параметъра съотносителност са съответно **корелат(+)** за рефлексивните по форма глаголни единици, които имат неререфлексивно по форма съответствие (например, *абонирам* / *абонирам се*, *вдъхновявам* / *вдъхновявам се*, *дърпам* / *дърпам се*; *abaissér* (спускам) / *s'abaissér* (спускам се), *libérer* (освобождавам) / *se libérer* (освобождавам се), *installer* (разполагам) / *s'installer* (разполагам се); *bezár* (затварям) / *bezárkózik* (затварям се), *visszahúz* (оттеглям) / *visszahúzódik* (оттеглям се), *megkeserít* (вгорчавам) / *megkeseredik* (вгорчавам се, огорчавам се)), и **корелат(-)** за тези, които нямат (например, *старая се*, *страхувам се*, *усмихвам се*; *s'efforcer* (старая се), *s'agenouiller* (коленича), *se moquer* (подигравам се); *bánkódik* (тъгувам), *kínlódik* (измъчвам се, страдам)).

Глаголните единици *корелат(+)* са многобройни и в трите езика и се подразделят в следните класове: глаголни единици със същинско рефлексивно значение; глаголни единици с реципрочно значение; глаголни единици, означаващи собствено движение; деакузативни глаголни единици; абсолютивни глаголни единици; антикаузативни глаголни единици; възвратен пасив.

3. Концептуални структури и моделиращи елементи

В Глава 3 са представени концептуалните структури, свързани с езиковите данни, както и съответните им моделиращи елементи. Подробно описание на релевантните за настоящата разработка моделиращи елементи в ЕПС е дадено в Приложение А към дисертацията. В изложението на Глава 3 са представени фундаментални концепти на модела и новите моделиращи елементи, които дефинирам като разширение на модела на ЕПС. Въвеждането на нови елементи е необходимо за построяването на семантичните дескриптори на изследваните тук глаголни единици.

3.1. Събитие

Според когнитивния подход семантичното представяне на даден глагол е равносилно на представяне на концепцията за *събитието*, което глаголът кодира. Събитието е концептуална единица в човешката когниция, „която включва в своите граници непрекъснатата корелация между поне някаква част от идентифициращата я качествена сфера и някаква част от темпоралния континуум (според нашите представи), т.е. от течението на времето. Тази корелация може да се основава на примитивна феноменологична представа, която да се определи като динамизъм – фундаментално свойство или принцип на дейност в света“ (Talmy 2000, с.215). Моделът на концептуалната единица събитие е СЪБИТИЙНАТА РАМКА, която съдържа градивните елементи на събитието.

3.2. Роли на участници

Една от спецификациите на КЛАСА УЧАСТНИЦИ е РОЛЯТА НА УЧАСТНИЦИ (PARTICIPANT ROLE), която се интерпретира като референция към семантична роля, изпълнявана от екземпляри на КЛАСА УЧАСТНИЦИ в рамките на събитието. Макар че РОЛЯТА НА УЧАСТНИЦИ се използва в семантичното описание в познатия лингвистичен смисъл, тя има специфични особености в системата на прилагания формализъм и в структурирането на дескрипторите. РОЛИТЕ НА УЧАСТНИЦИ като част от спецификациите на КЛАСОВЕТЕ УЧАСТНИЦИ са свързани с *ролеви* КЛАСОВЕ, които дефинират прототипни семантични роли. Изразът за роля е име на *ролеви* КЛАС, който се различава от „нормалните“ КЛАСОВЕ, тъй като представлява съвкупност от АТРИБУТИ (ATTRIBUTE clusters), които дефинират понятие за роля, а не множество от възможни ОБЕКТИ като екземпляри на КЛАСА. В определен смисъл *ролевите* КЛАСОВЕ представляват макроси на съвкупности от прототипни АТРИБУТИ, т.е. имената на роли са удобни мнемонични названия на концептуални конфигурации. Тези концептуални конфигурации се състоят от релационни семантични признаци (relational semantic features), т.е. признаци, които характеризират релацията между участниците и динамичното ядро на събитието.

Всеизвестно е, че в лингвистичната литература няма единство по въпроса за системата от семантични роли, като съществуват големи различия в становищата относно броя и вида на семантичните роли. В ЕПС е дадена възможността за дефиниране на семантични роли, тъй като те, въпреки своята неустановеност, представляват дълбоко вкоренено средство в лингвистичните описания. Също така специфичният подход в ЕПС към семантичните роли ги превръща в подходящо допълнение на моделите. РОЛИТЕ НА УЧАСТНИЦИ се дефинират като съвкупности от свойства. В ЕПС се предлага набор от възможни РОЛИ НА УЧАСТНИЦИ, но той не е фиксиран – потребителят има възможността да дефинира и адаптира необходимите за конкретното описание роли. Важна особеност на системата от семантични роли в ЕПС е прилагането на подхода на Даути (Dowty 1991), според който типовете семантични роли не са дискретни категории, а образуват континуум от концепти и всъщност се наблюдава степенуване на членството на обектите към даден тип роля. Оттук се

формулират и понятията за макророли (macroroles) – обобщени типове семантични роли, които обхващат няколко по-специфични роли, разположени в близост една до друга върху континуума от тематични релации на аргументите на даден предикат. Обобщеният тип АГЕНС се нарича *действащ*, а обобщеният тип ПАЦИЕНС се нарича *търпящ* (undergoer).

За целите на настоящата разработка дефинирам система от РОЛИ НА УЧАСТНИЦИ, които са съобразени с принципите на ЕПС, но са резултат от моите собствени наблюдения и решения относно необходимите за моделите роли и техните значения. Като се има предвид, че в ЕПС РОЛИТЕ НА УЧАСТНИЦИ са съвкупности от релационни признаци, които при това се разграничават от вътрешно-присъщите признаци на участниците, може да се дефинира сравнително стройна система от *ролеви* КЛАСОВЕ. Използва се също така възможността за дефиниране на семантичните роли в йерархии.

Един глобален лингвистичен поглед върху основните типове залогови отношения и диатезни редувания в аргументната структура на глаголните предикати показва, че се очертават три абстрактни *ролеви* КЛАСА. Първият *ролеви* КЛАС е дефиниран във формализма на ЕПС като *Причинител* с елементи *Агенс* и *Ефектор*. Аналогично на тази йерархия въвеждам абстрактния *ролеви* КЛАС *Причинен* с подкласове *Пациенс* и *Тема*. Абстрактната роля *Причинен* е обобщение на РОЛЯТА НА УЧАСТНИК, търпящ някакво въздействие в събитието, независимо дали това е преход на състоянието му или просто локализиране или преместване. Както абстрактният *ролеви* КЛАС *Причинител*, така и *Причинен* не може да има свои екземпляри, а трябва първо да получи като стойност един от *ролевите* КЛАСОВЕ *Пациенс* или *Тема*. Третият абстрактен *ролеви* КЛАС е *Медиум* с подкласове *Експериенцер* и *Повлиян*. Въвеждам абстрактната роля *Медиум* като обобщение на РОЛЯТА НА УЧАСТНИК, който се намира в средно положение между един активен деятел (прототипен АГЕНС) и един пасивен участник, търпящ въздействие (прототипен ПАЦИЕНС). Затова участник с ролята *Експериенцер* в едни събития може да изпълнява макроролята *Действащ*, а в други събития макроролята *Търпящ*. Най-общо казано, абстрактната роля *Медиум* и нейните елементи *Експериенцер* и *Повлиян* са част от дескриптори на събития, които се възприемат като случващи се от само себе си.

3.3. Семантични релации между участниците

Предварително зададеното в ЕПС ОГРАНИЧЕНИЕ *потенциаленРефлексив* използвам при моделирането на глаголни единици със същинско рефлексивно значение. ОГРАНИЧЕНИЕТО *потенциаленРефлексив* означава, че съществуват ОБЕКТИ-УЧАСТНИЦИ, които могат да бъдат екземпляри едновременно на всички (на практика на два) КЛАСОВЕ УЧАСТНИЦИ, които се намират в обсега на ОГРАНИЧЕНИЕТО, т.е. в рамките на една и съща диаграма се моделират както рефлексивни, така и нереплексивни варианти на глагола. По този начин прилагам ОГРАНИЧЕНИЕТО *потенциаленРефлексив* за създаване на генерализиран модел на преходни глаголи, които могат да участват и в същински рефлексивни композиции, например, (*Тя се мие*, фр. *Elle se lave*, унр. *Ő mosakodik*).

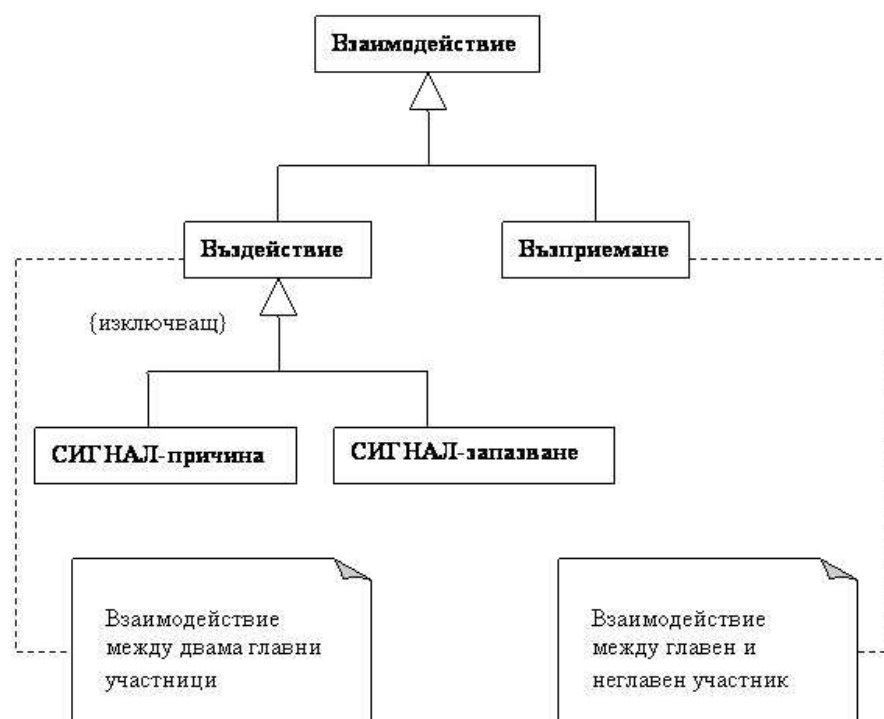
В реципрочната композиция (например, *измъчваме се (един друг)*, *хвалим се (един друг)*, *държим се (един друг)*) двама главни участници си взаимодействат така, че извършвайки едно и също действие, всеки един от тях въздейства върху другия участник и същевременно търпи въздействие от другия участник, т.е. взаимодействието е симетрично. За да изразя този тип взаимодействие между участниците, дефинирам нов тип ОГРАНИЧЕНИЕ – ОГРАНИЧЕНИЕТО Реципрочност. То се отнася до КЛАСОВЕТЕ УЧАСТНИЦИ и означава, че всички ОБЕКТИ-УЧАСТНИЦИ, които са екземпляри на единия КЛАС УЧАСТНИЦИ са същевременно екземпляри и на другия КЛАС УЧАСТНИЦИ.

3.4. Типове събития и пакети от моделиращи елементи

В Глава 3 са разгледани типове взаимодействия между обекти, както и типове събития, свързани с поведението на един обект. Дефинирани са концепти на прототипни събития и са конструирани съответните схеми в термините на ЕПС. В настоящата разработка е разширена предложената от А. Шели теоретично обоснована класификация на събития (с акцент върху динамичната структура на събитията) с модели, получени въз основа на прилагането на формализма на ЕПС върху множество от реални езикови данни, като се разглеждат варианти на събития с включени участници и асоциативни релации между тях.

Поради многозначността на описваните глаголни единици в българския език и съответствията им във френски и унгарски език, се оформят множество концептуални структури, представени чрез комбинации от моделиращи елементи на ЕПС. Изнесените на преден план субектно-обектни отношения се моделират от два основни типа СЪБИТИЙНИ РАМКИ – СЪБИТИЙНА РАМКА с двама участници и СЪБИТИЙНА РАМКА с един участник. Характерът на взаимодействието между двама участници или поведението на един участник предопределят наличието на различните типове събития. На Фиг. 2 е дадена таксономията на концептуалните структури и съответните СЪБИТИЙНИ РАМКИ, при които двама участници си взаимодействат.

На най-високо ниво на генерализация се очертават два типа взаимодействие между двамата участници: *въздействие* на даден участник върху друг участник; *възприемане* от даден участник на друг участник. Типът *въздействие* на един участник върху друг участник се подразделя на два взаимно изключващи се подтипа, което е посочено от ОГРАНИЧЕНИЕТО *изключващ*. Първият подтип *въздействие* е този, при който даден участник въздейства върху друг участник така, че вторият участник претърпява промяна в състоянието си. Въздействието, както и промяната в състоянието, се разглеждат в най-общ смисъл – физическо въздействие и промяна (*изяждам, изсушавам*), в това число и преместване (*поставям, придвижвам*); психическо въздействие и промяна (*убеждавам, информирам*), абстрактно въздействие и промяна, свързани с различни дейности в обществото (*абонирам, приобщавам, интегрирам*). При втория подтип *въздействие* на даден участник върху друг участник не се концептуализира промяна в състоянието на участника, търпящ въздействие. Това, което се концептуализира е стремежът на първия главен участник да запази състоянието, в което се намира вторият участник. Този тип модел на събитие до-



Фиг. 2. Типове събития с *взаимодействие* на двама участници

бавям към класификацията на концептуалните структури, предложена от А. Шели, като за целта въвеждам и съответни елементи в метамодела на ЕПС. В представения от А. Шели метамодел има дефиниран един единствен СИГНАЛ – това е СИГНАЛ-причина, който задължително предизвиква ПРЕХОД в СЪСТОЯНИЕТО на втория главен участник, като се предвиждат варианти, свързани с различни комбинации на типове СЪСТОЯНИЯ, участващи в механизмите на СЪСТОЯНИЯТА, съответстващи на участниците. Прилагането на моделиращия апарат върху реалните езикови данни показва, че се концептуализират значителен брой събития, в които са налице двама главни участници, но първият участник се стреми така да въздейства върху втория участник, че той да остане в СЪСТОЯНИЕТО, в което се намира, например, *пазя, съхранявам, задържам, държа книгите в горния шкаф*. Такава концептуална структура налага въвеждането на още един тип СИГНАЛ – СИГНАЛ-запазване. Приемането на такъв СИГНАЛ от втория главен участник предизвиква запазване, оставане в СЪСТОЯНИЕТО, в което се намира. При взаимодействието между двама участници от типа *възприемане* от даден участник на друг участник, събитието представлява вътрешно преживяване на единия участник, свързано с втория участник, например, даден участник възприема друг участник (*виждам, чувам, пипам, броя*), или усеща, чувства нещо спрямо друг участник (*мразя, обичам*), или изразява отношение, оценка към друг участник (*набеждам, виня, критикувам*). В концептуалната структура на *възприемането* има един главен участник - това е възприеманият и един неглавен участник - възприеманият.

Вторият основен тип събития са тези с един участник. Участникът се намира в дадено СЪСТОЯНИЕ, а може да претърпи и ПРЕХОД от едно СЪСТОЯНИЕ в друго. Класификацията на събитията с един участник зависи от наличието или отсъствието на ПРЕХОД на СЪСТОЯНИЯТА, както и от възможните комбинации между изходни и крайни СЪСТОЯНИЯ при ПРЕХОД.

4. Семантично интерпретирани глаголни структури

В Глава 4 е представена базата от знания **Семивест**. Глаголните единици са подредени в *таблицы с данни*, предназначени за свързване в една реляционна база данни. Първият тип таблица съдържа поле с българските глаголи от даден клас и поле на първичен ключ, който идентифицира таблицата в структурата на базата данни като таблица - родител и дава възможност за установяване на релации с други, дъщерни, таблици. Таблицата съдържа и поле на идентификаторите на събитийните рамки - шаблони, съответстващи на глаголите от таблицата. Вторият тип таблица съдържа поле с еквивалентни глаголи на френски или на унгарски език, чийто ключ съвпада с първичния ключ на таблицата - родител с българските глаголи. Типът релация е *едно към много*, т.е. на един български глагол съответства един или повече от един глагол съответно на френски или на унгарски език, но на даден френски или унгарски глагол съответства точно един български глагол. Допълнителните еквиваленти на френски или унгарски са представени в отделни таблици. За удобство дъщерните таблици включват повтаряща се информация от таблицата - родител, а

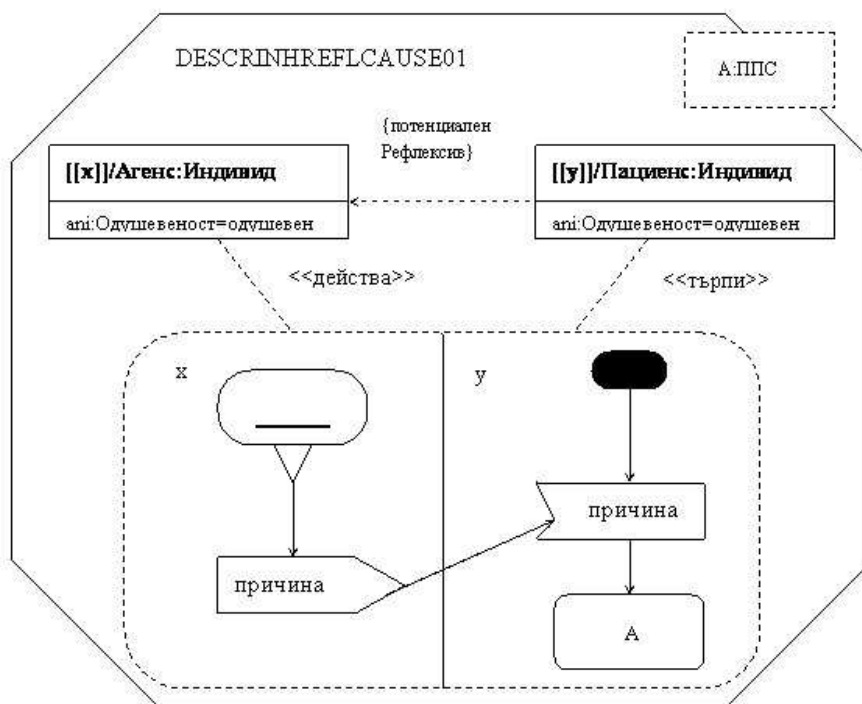
именно, полето на идентификаторите на събитийните рамки - шаблони, съответстващи на глаголите от таблицата. За всеки клас глаголни единици са дадени също така диаграмите на СЪБИТИЙНИТЕ РАМКИ ШАБЛОНИ, които се асоциират с глаголите, посочени в съответния клас. Дадени са и примери на СЪБИТИЙНИ РАМКИ, в които всички параметри са получили стойности, т.е. това са специфицирани СЪБИТИЙНИ РАМКИ, описващи значенията на дадени глаголи.

4.1. Дескриптори на глаголни единици със същинско рефлексивно значение

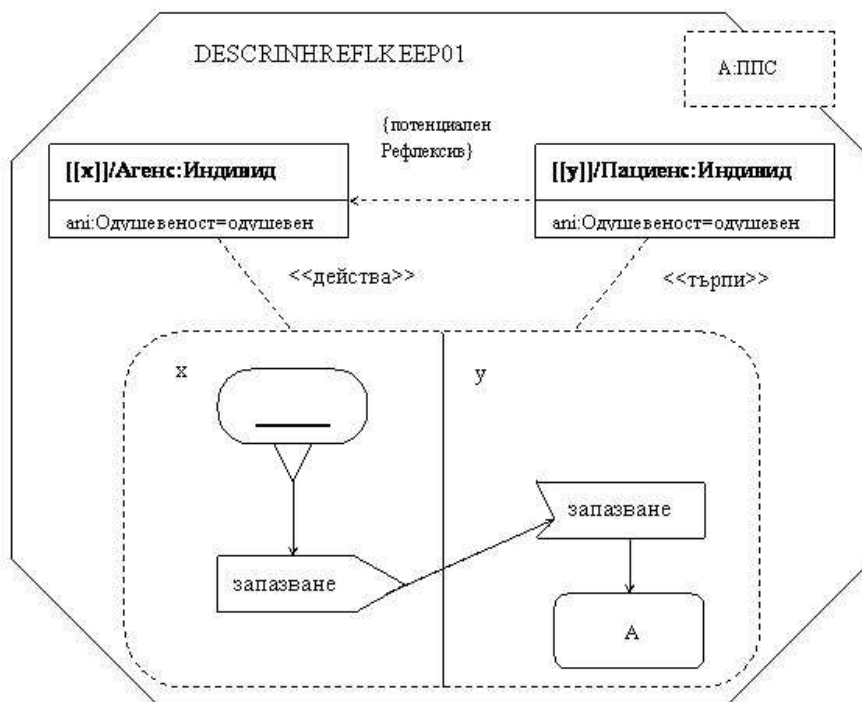
Глаголните единици със същинско рефлексивно значение се разглеждат в единство с нереклексивните им транзитивни корелати, при които двама участници си взаимодействат. Характерът на действието, изразено чрез нереклексивните корелати, се запазва при съответните им рефлексивни глаголи и това се изразява в модела чрез наличието на едни и същи елементи в динамичното ядро – едни и същи механизми на СЪСТОЯНИЯТА. Разликата е в денотацията на участниците. При нереклексивните корелати референтите на участниците са различни и това се изразява в модела чрез инстанцирането на КЛАСОВЕТЕ УЧАСТНИЦИ с различни ОБЕКТИ. При същинските рефлексивни глаголи референтът на участниците е един и същ, т.е. действието произхожда от даден участник и се извършва върху самия него по същия начин както това става върху втория участник при нереклексивните корелати. В модела такова положение се изразява чрез инстанцирането на КЛАСОВЕТЕ УЧАСТНИЦИ с един и същ ОБЕКТ. В лексикона значителен брой транзитивни глаголи притежават потенциал за образуване на същински рефлексивни съответствия, чиито израз на морфосинтактично ниво е както под формата на лексеми (например *абонирам* - *абонирам се*), така и под формата на синтактична структура с пълнозначно местоимение (например, *заставям себе си*, *убивам себе си*). Следователно същинското рефлексивно значение се проявява като регулярно изменение на значението на глаголите и затова може да се представи чрез генерализирана СЪБИТИЙНА РАМКА, описваща както транзитивно нереклексивно значение, така и същинско рефлексивно значение на даден глагол. Моделирането на същинските рефлексивни глаголни единици в единство с нереклексивните им корелати означава, че за същинските рефлексиви е валидна класификацията на типовете събития с *взаимодействие* на двама участници. Това са типове събития със СИГНАЛ- *причина*, със СИГНАЛ- *запазване* и тип събитие *възприемане*. На Фигури 3, 4 и 5 са дадени част от диаграмите на СЪБИТИЙНИТЕ РАМКИ - ШАБЛОНИ, които описват нереклексивно значение и потенциално същинско рефлексивно значение. На Фигура 6 е даден пример за специфицирана СЪБИТИЙНА РАМКА. В Таблицы 1, 2 и 3 са дадени част от включените в базата от данни глаголни единици със същинско рефлексивно значение на български, френски и унгарски език.

4.2. Дескриптори на глаголни единици с реципрочно значение

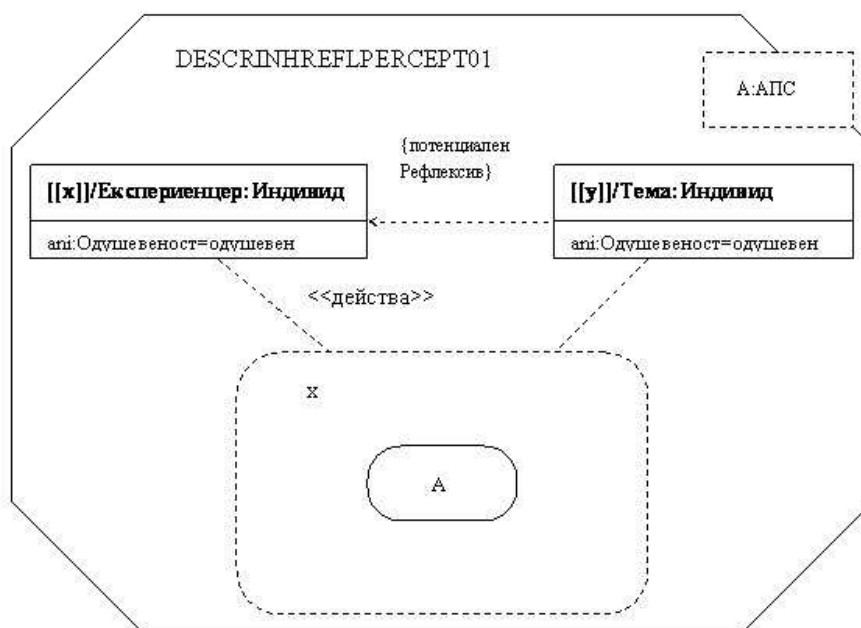
СЪБИТИЙНИТЕ РАМКИ - ШАБЛОНИ, които съответстват на глаголни единици с реципрочно значение, включват двама участници, които си взаимодействат така, че



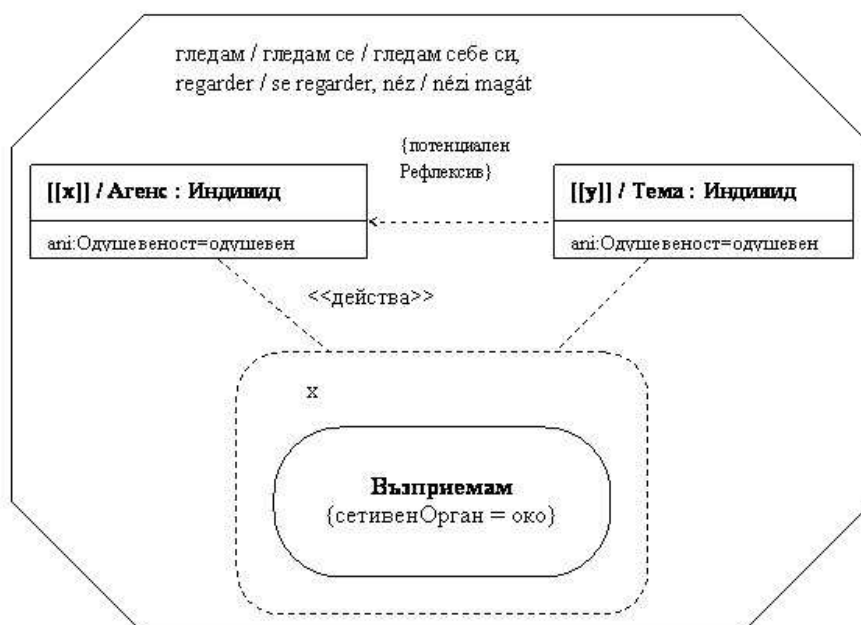
Фиг. 3. Рефлексивно значение със СИГНАЛ- причина: вариант 1



Фиг. 4. Рефлексивно значение със СИГНАЛ- запазване: вариант 1



Фиг. 5. Рефлексивно значение - тип събитие *възприемане*



Фиг. 6. Дескриптор на *гледам/гледам се (себе си), regarder/se regarder, néz/nézi magát*

СъщРефлБгИд	КорелатБг	СъщРефлБг	СъщРефлСъБРамШаблИд
INHREFLBG0001	беда	беда се/себе си	DESCRINHREFLPERCEPT01
INHREFLBG0002	беся	беся се/себе си	DESCRINHREFLCAUSE01
INHREFLBG0003	брана	брана се/себе си	DESCRINHREFLKEEP02
INHREFLBG0004	броя	броя се/себе си	DESCRINHREFLPERCEPT01
INHREFLBG0005	виждам	виждам се/себе си	DESCRINHREFLPERCEPT01
INHREFLBG0006	вина	вина се/себе си	DESCRINHREFLPERCEPT01
INHREFLBG0007	вредя	вредя си/на себе си	DESCRINHREFLCAUSE01
INHREFLBG0008	гледам	гледам се/себе си	DESCRINHREFLPERCEPT01
INHREFLBG0009	диря	диря се/себе си	DESCRINHREFLKEEP02
INHREFLBG0010	докарвам	докарвам се/себе си	DESCRINHREFLCAUSE01
INHREFLBG0011	забелязвам	забелязвам се/себе си	DESCRINHREFLPERCEPT01
INHREFLBG0012	задължавам	задължавам се/себе си	DESCRINHREFLCAUSE01
INHREFLBG0013	запасявам	запасявам се/себе си	DESCRINHREFLCAUSE01
INHREFLBG0014	зарязвам	зарязвам се/себе си	DESCRINHREFLCAUSE01
INHREFLBG0015	извинявам	извинявам се/себе си	DESCRINHREFLKEEP01
INHREFLBG0016	издържам	издържам се/себе си	DESCRINHREFLKEEP01
INHREFLBG0017	изключвам	изключвам се/себе си	DESCRINHREFLKEEP01
INHREFLBG0018	измъчвам	измъчвам се/себе си	DESCRINHREFLCAUSE01
INHREFLBG0019	изследвам	изследвам се/себе си	DESCRINHREFLPERCEPT01
INHREFLBG0020	изтъквам	изтъквам се/себе си	DESCRINHREFLKEEP01
INHREFLBG0021	изхранвам	изхранвам се/себе си	DESCRINHREFLKEEP01
INHREFLBG0022	капя	капя се/себе си	DESCRINHREFLCAUSE02
INHREFLBG0023	коря	коря се/себе си	DESCRINHREFLPERCEPT01
INHREFLBG0024	крася	крася се/себе си	DESCRINHREFLCAUSE01
INHREFLBG0025	лишавам	лишавам се/себе си	DESCRINHREFLCAUSE03
INHREFLBG0026	мия	мия се/себе си	DESCRINHREFLCAUSE01
INHREFLBG0027	мразя	мразя се/себе си	DESCRINHREFLPERCEPT01
INHREFLBG0028	наблюдавам	наблюдавам се/себе си	DESCRINHREFLPERCEPT01
INHREFLBG0029	навличам	навличам се/себе си	DESCRINHREFLCAUSE02
INHREFLBG0030	наказвам	наказвам се/себе си	DESCRINHREFLCAUSE01
INHREFLBG0031	направлявам	направлявам се/себе си	DESCRINHREFLCAUSE01
INHREFLBG0032	наранявам	наранявам се/себе си	DESCRINHREFLCAUSE01
INHREFLBG0033	насърчавам	насърчавам се/себе си	DESCRINHREFLKEEP02
INHREFLBG0034	нахранвам	нахранвам се/себе си	DESCRINHREFLCAUSE02
INHREFLBG0035	обвинявам	обвинявам се/себе си	DESCRINHREFLPERCEPT01
INHREFLBG0036	обвързвам	обвързвам се/себе си	DESCRINHREFLCAUSE01
INHREFLBG0037	обесвам	обесвам се/себе си	DESCRINHREFLCAUSE01
INHREFLBG0038	обличам	обличам се/себе си	DESCRINHREFLCAUSE02
INHREFLBG0039	оборудвам	оборудвам се/себе си	DESCRINHREFLCAUSE01
INHREFLBG0040	образовам	образовам се/себе си	DESCRINHREFLCAUSE01
INHREFLBG0041	обричам	обричам се/себе си	DESCRINHREFLCAUSE01

Таблица 1. Глаголни единици на български език със същинско рефлексивно значение: част 1

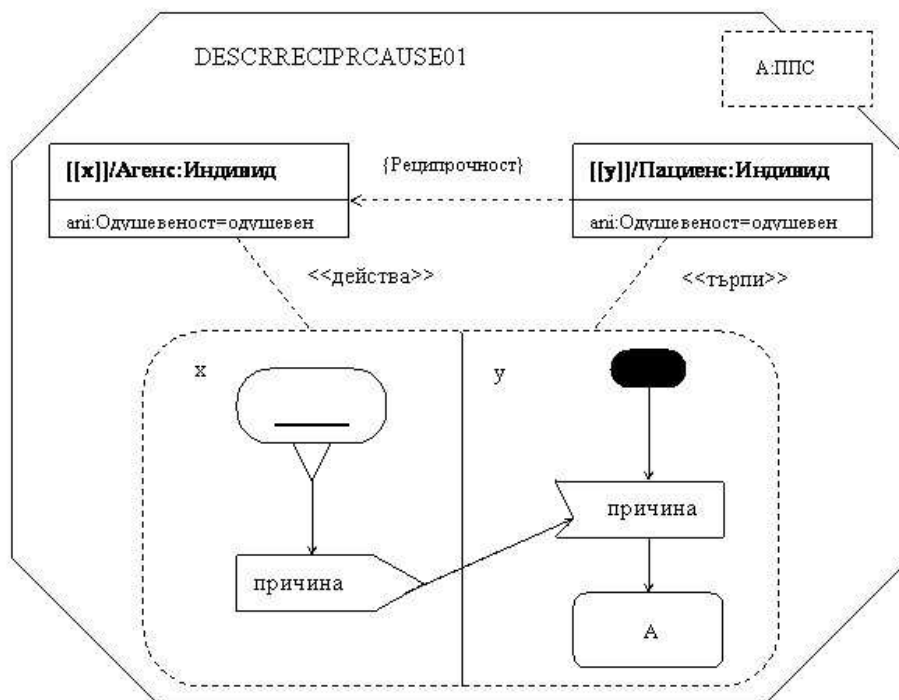
СъщРефлБгИд	КорелатФр01	СъщРефлФр02	СъщРефлСъбРамШаблИд
INHREFLBG0001	accuser	s'accuser	DESCRINHREFLPERCEPT01
INHREFLBG0002	pendre	se pendre	DESCRINHREFLCAUSE01
INHREFLBG0003	défendre	se défendre	DESCRINHREFLKEEP02
INHREFLBG0004	compter	se compter	DESCRINHREFLPERCEPT01
INHREFLBG0005	voir	se voir	DESCRINHREFLPERCEPT01
INHREFLBG0006	accuser	s'accuser	DESCRINHREFLPERCEPT01
INHREFLBG0007	faire du tort à	se faire du tort	DESCRINHREFLCAUSE01
INHREFLBG0008	regarder	se regarder	DESCRINHREFLPERCEPT01
INHREFLBG0009	chercher	se chercher	DESCRINHREFLKEEP01
INHREFLBG0010	parer	se parer	DESCRINHREFLCAUSE01
INHREFLBG0011	apercevoir	s'apercevoir	DESCRINHREFLPERCEPT01
INHREFLBG0012	obliger	s'obliger	DESCRINHREFLCAUSE01
INHREFLBG0013	approvisionner	s'approvisionner	DESCRINHREFLCAUSE01
INHREFLBG0014	négliger	se négliger	DESCRINHREFLCAUSE01
INHREFLBG0015	excuser	s'excuser	DESCRINHREFLKEEP01
INHREFLBG0016	entretenir	s'assurer	DESCRINHREFLKEEP01
INHREFLBG0017	excepter	s'excepter	DESCRINHREFLKEEP01
INHREFLBG0018	torturer	se torturer	DESCRINHREFLCAUSE01
INHREFLBG0019	examiner	s'examiner	DESCRINHREFLPERCEPT01
INHREFLBG0020	faire valoir	se faire valoir	DESCRINHREFLKEEP01
INHREFLBG0021	entretenir	s'assurer	DESCRINHREFLKEEP01
INHREFLBG0022	tacher	se tacher	DESCRINHREFLCAUSE02
INHREFLBG0023	blâmer	se blâmer	DESCRINHREFLPERCEPT01
INHREFLBG0024	parer	se parer	DESCRINHREFLCAUSE01
INHREFLBG0025	priver	se priver	DESCRINHREFLCAUSE03
INHREFLBG0026	laver	se laver	DESCRINHREFLCAUSE01
INHREFLBG0027	haïr	se haïr	DESCRINHREFLPERCEPT01
INHREFLBG0028	observer	s'observer	DESCRINHREFLPERCEPT01
INHREFLBG0029	emmitoufler	s'emmitoufler	DESCRINHREFLCAUSE02
INHREFLBG0030	punir	se punir	DESCRINHREFLCAUSE01
INHREFLBG0031	diriger	se diriger	DESCRINHREFLCAUSE01
INHREFLBG0032	blesser	se blesser	DESCRINHREFLCAUSE01
INHREFLBG0033	encourager	s'encourager	DESCRINHREFLKEEP02
INHREFLBG0034	rassasier	se rassasier	DESCRINHREFLCAUSE02
INHREFLBG0035	accuser	s'accuser	DESCRINHREFLPERCEPT01
INHREFLBG0036	lier	se lier	DESCRINHREFLCAUSE01
INHREFLBG0037	pendre	se pendre	DESCRINHREFLCAUSE01
INHREFLBG0038	vêtir	se vêtir	DESCRINHREFLCAUSE02
INHREFLBG0039	outiller	s'outiller	DESCRINHREFLCAUSE01
INHREFLBG0040	instruire	s'instruire	DESCRINHREFLCAUSE01
INHREFLBG0041	vouer	se vouer	DESCRINHREFLCAUSE01

Таблица 2. Еквиваленти 01 на френски език на глаголни единици на български език със същинско рефлексивно значение: част 1

СъщРефлБгИд	КорелатУнг01	СъщРефлУнг01	СъщРефлСъбРамШаблИд
INHREFLBG0001	vádol	vádolja magát	DESCRINHERREFLPERCEPT01
INHREFLBG0002	akaszt	akasztja magát	DESCRINHERREFLCAUSE01
INHREFLBG0003	véd	védekezik	DESCRINHERREFLKEEP02
INHREFLBG0004	számít	számítja magát	DESCRINHERREFLPERCEPT01
INHREFLBG0005	lát	látja magát	DESCRINHERREFLPERCEPT01
INHREFLBG0006	hibáztat	hibáztatja magát	DESCRINHERREFLPERCEPT01
INHREFLBG0007	árt vkinek	árt magának	DESCRINHERREFLCAUSE01
INHREFLBG0008	néz	nézi magát	DESCRINHERREFLPERCEPT01
INHREFLBG0009	keres	keresi magát	DESCRINHERREFLKEEP01
INHREFLBG0010	kiöltöztet	kiöltözik	DESCRINHERREFLCAUSE01
INHREFLBG0011	észrevesz	észreveszi magát	DESCRINHERREFLPERCEPT01
INHREFLBG0012	kötelez	kötelezi magát	DESCRINHERREFLCAUSE01
INHREFLBG0013	felszerel	felszereli magát	DESCRINHERREFLCAUSE01
INHREFLBG0014	elhanyagol	elhanyagolja magát	DESCRINHERREFLCAUSE01
INHREFLBG0015	megbocsát	megbocsát magának	DESCRINHERREFLKEEP01
INHREFLBG0016	eltart	eltartja magát	DESCRINHERREFLKEEP01
INHREFLBG0017	kirekeszt	kirekeszt magát	DESCRINHERREFLKEEP01
INHREFLBG0018	kínoz	kínozza magát	DESCRINHERREFLCAUSE01
INHREFLBG0019	vizsgál	vizsgálja magát	DESCRINHERREFLPERCEPT01
INHREFLBG0020	magasztal	magasztalja magát	DESCRINHERREFLKEEP01
INHREFLBG0021	eltart	eltartja magát	DESCRINHERREFLKEEP01
INHREFLBG0022	cseppent	cseppent magára	DESCRINHERREFLCAUSE02
INHREFLBG0023	korhol	korholja magát	DESCRINHERREFLPERCEPT01
INHREFLBG0024	ékesít	ékesíti magát	DESCRINHERREFLCAUSE01
INHREFLBG0025	megfoszt	megfosztja magát	DESCRINHERREFLCAUSE03
INHREFLBG0026	mos	mossa magát	DESCRINHERREFLCAUSE01
INHREFLBG0027	gyűlöl	gyűlöli magát	DESCRINHERREFLPERCEPT01
INHREFLBG0028	figyel	figyeli magát	DESCRINHERREFLPERCEPT01
INHREFLBG0029	felöltöztet melegen	felöltözik melegen	DESCRINHERREFLCAUSE02
INHREFLBG0030	megbüntet	megbünteti magát	DESCRINHERREFLCAUSE01
INHREFLBG0031	irányít	irányul	DESCRINHERREFLCAUSE01
INHREFLBG0032	megsebesít	megsebesíti magát	DESCRINHERREFLCAUSE01
INHREFLBG0033	buzdít	buzdítja magát	DESCRINHERREFLKEEP02
INHREFLBG0034	megetet	megeteti magát	DESCRINHERREFLCAUSE02
INHREFLBG0035	vádol	vádolja magát	DESCRINHERREFLPERCEPT01
INHREFLBG0036	elkötelez	elkötelezi magát	DESCRINHERREFLCAUSE01
INHREFLBG0037	felakaszt	felakasztja magát	DESCRINHERREFLCAUSE01
INHREFLBG0038	öltöztet	öltözködik	DESCRINHERREFLCAUSE02
INHREFLBG0039	felszerel	felszereli magát	DESCRINHERREFLCAUSE01
INHREFLBG0040	oktat	oktatja magát	DESCRINHERREFLCAUSE01
INHREFLBG0041	szentel	szenteli magát	DESCRINHERREFLCAUSE01

Таблица 3. Еквиваленти 01 на унгарски език на глаголни единици на български език със същинско рефлексивно значение: част 1

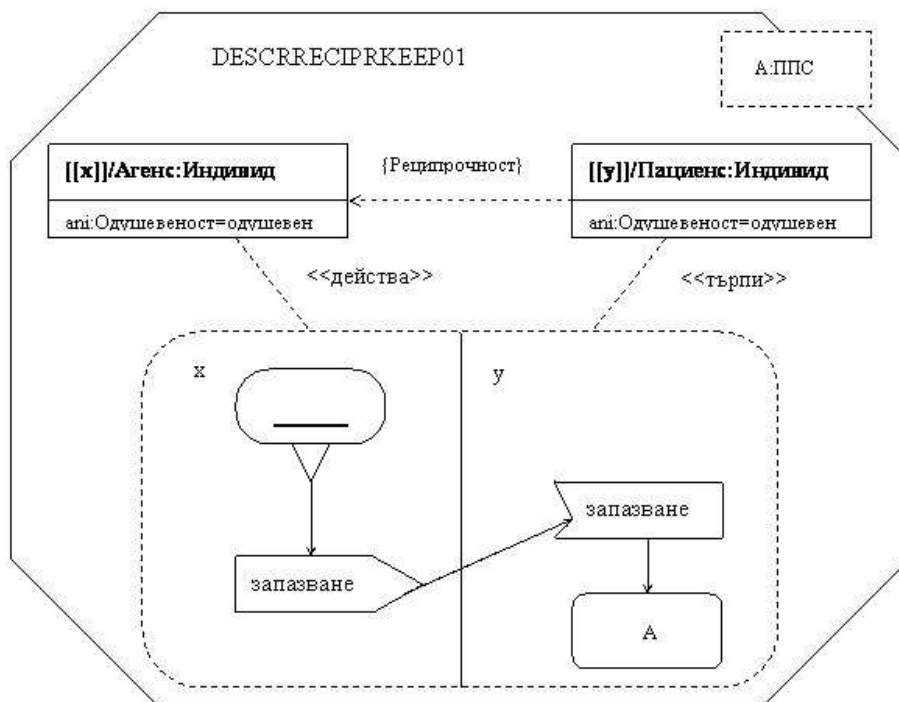
извършвайки едно и също действие, всеки един от тях въздейства върху другия участник и същевременно търпи въздействие от другия участник, т.е. взаимодействието е симетрично. За да се изрази този тип взаимодействие между участниците, бе дефинирано ново ОГРАНИЧЕНИЕ – ОГРАНИЧЕНИЕТО Реципрочност. За реципрочните глаголни единици са валидни типове събития със СИГНАЛ- *причина*, със СИГНАЛ- *запазване* и тип събитие *възприемане*. На Фигури 7, 8 и 9 са дадени част от диаграмите на СЪБИТИЙНИТЕ РАМКИ - ШАБЛОНИ, които описват глаголни единици с реципрочно значение.



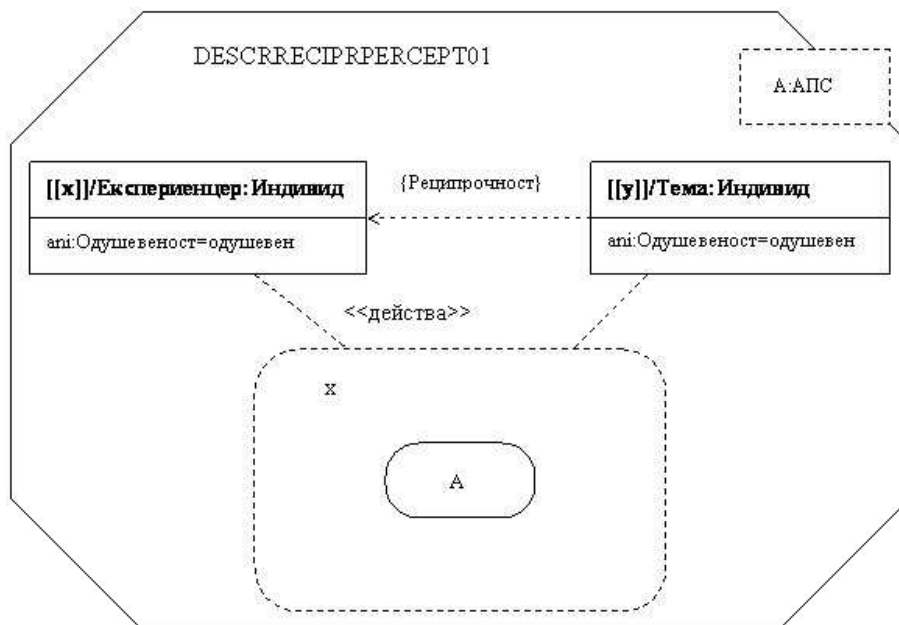
Фиг. 7. Реципрочно значение със СИГНАЛ- причина: вариант 1

4.3. Дескриптори на глаголни единици, означаващи собствено движение

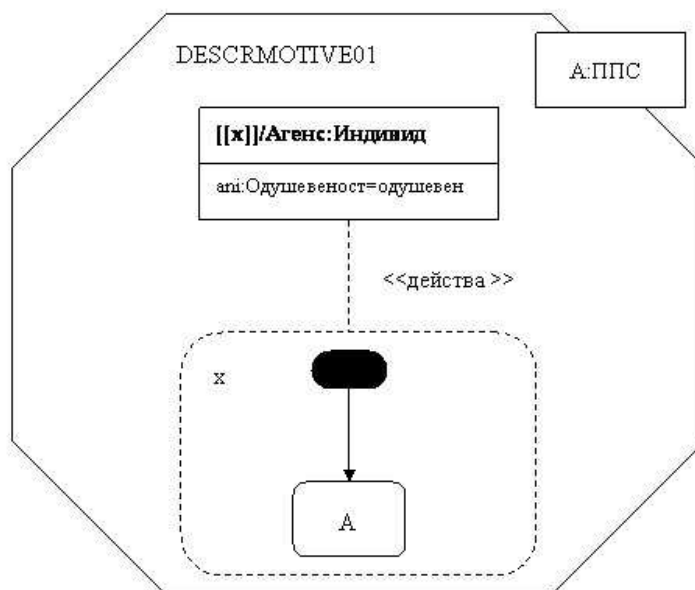
Глаголните единици, означаващи собствено движение се описват от СЪБИТИЙНИ РАМКИ с един участник в ролята на Агeнс, който, в най-общия смисъл, променя своето собствено място или положение в пространството, например, *хвърлям се* – фр. *se jeter, se lancer* – унг. *veti magát*; *дрънвам се* – фр. *se tirer* - унг. *elhúzódik*; *издигам се* – фр. *s'élever* – унг. *felemelkedik*. На Фиг. 10 е даден пример на СЪБИТИЙНА РАМКА - ШАБЛОН, която представя глаголните единици, свързани със собствено значение, включени понастоящем в **Семиналест**.



Фиг. 8. Реципрочно значение със СИГНАЛ- запазване: вариант 1



Фиг. 9. Реципрочно значение - тип събитие *възприемане*

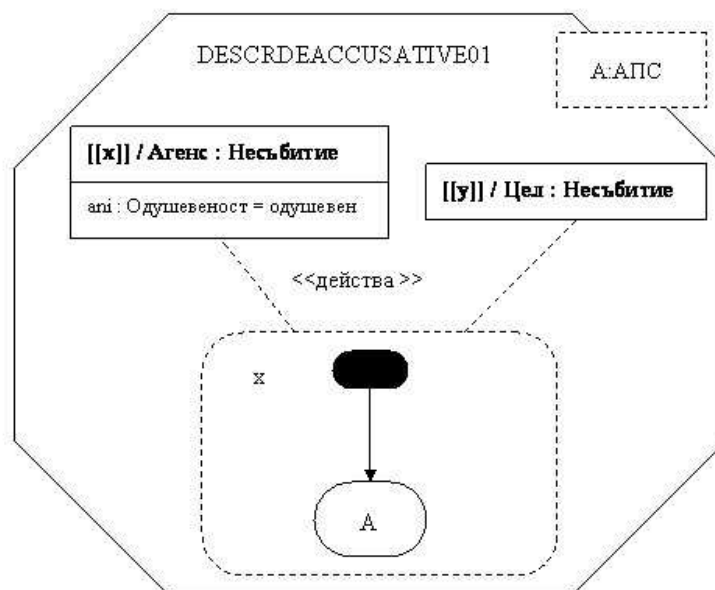
Фиг. 10. Значение *собствено движение*: вариант 1

4.4. Дескриптори на деакузативни глаголни единици

Броят на деакузативните глаголни единици в българския език е малък. Еквивалентите им във френския език само в някои случаи са също деакузативни глаголни единици, например, *държа* / *държа се за* – *tenir* / *se tenir à*, *срещам* / *срещам се с* – *rencontrer* / *se rencontrer avec*, *отказвам* / *отказвам се от* – *refuser* / *se refuser à*. В унгарския език еквивалентите, които могат да се определят като деакузативни, са още по-малобройни, например, *опитвам* / *опитвам се* – *próbál* / *próbálkozik*, *държа* / *държа се за* – *fog* / *fogódzik*. При деакузативните глаголни единици степента на промяна на лексикалното значение на глагола спрямо транзитивния му корелат е различна, но се наблюдава насоченост на действието към Субекта, което се представя чрез СЪБИТИЙНА РАМКА - ШАБЛОН с един главен участник. На Фиг. 11 е даден пример на СЪБИТИЙНА РАМКА - ШАБЛОН, моделираща значенията на глаголните единици от този клас.

4.5. Дескриптори на абсолютивни глаголни единици

Абсолютивните глаголни единици се описват от СЪБИТИЙНИ РАМКИ с един участник, който в преобладаващите случаи е в ролята на Агенс, например, *блъскам се* – фр. *se pousser*, *se bousculer* – унг. *lökösödik*, *tolakodik*; *овладявам се* – фр. *se maîtriser*, *se contenir* – унг. *uralkodik magán*; *упражнявам се* – фр. *s'exercer* – унг. *gyakorolja magát*. В някои случаи СЪБИТИЙНАТА РАМКА съдържа абстрактния ролеви клас *Причинител*, който може да получи стойност Агенс или Ефектор, например, *вливам се* – фр. *se jeter* – унг. *ömlik*, *beömlik*; *подготвям се* – фр. *se préparer* – унг. *készül*,



Фиг. 11. Дескриптор на деакузативни глаголни единици: вариант 1

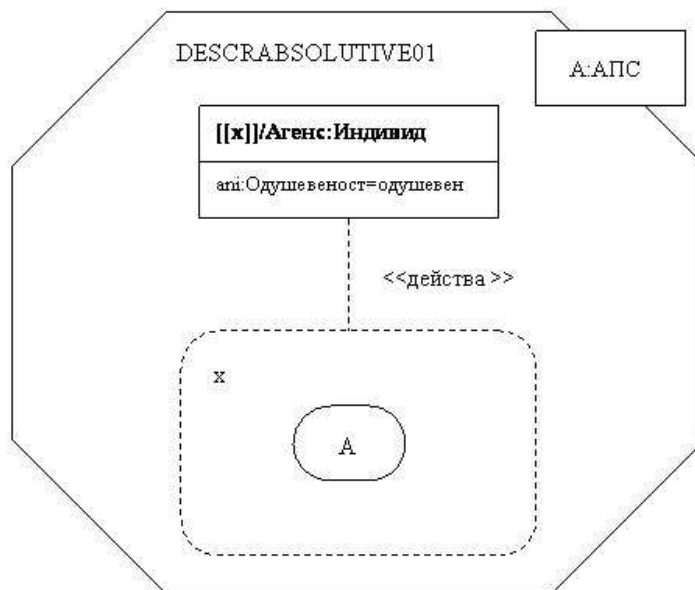
felkészül. Възможна е и ролята Експериенцер, например, *чувствам се* – фр. *se sentir* – унг. *érzi magát.* Характерът на действията, обозначени с абсолютни глаголни единици, предполага в повечето случаи одушевен участник. На Фигура 12 е даден пример на СЪБИТИЙНА РАМКА - ШАБЛОН, съответстваща на абсолютни глаголни единици, включени понастоящем в Семинвест.

4.6. Дескриптори на антикаузативни глаголни единици

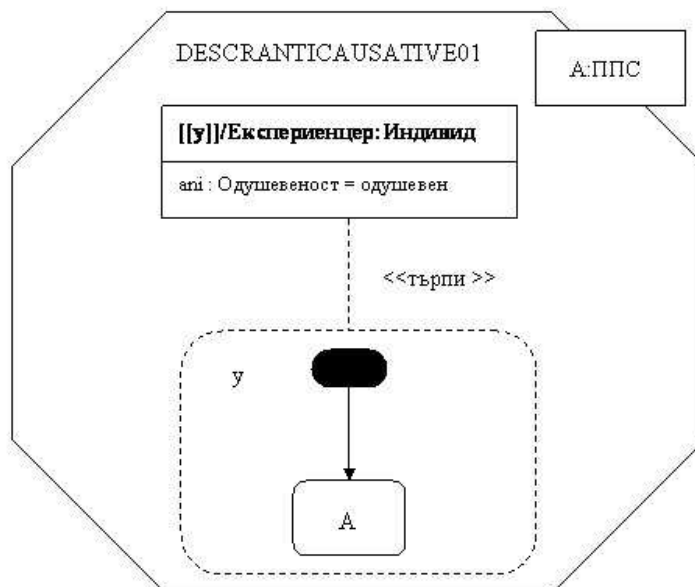
При антикаузативните глаголни единици действието, най-общо казано, се възприема като „действие, ставащо от само себе си“. Акцентът е върху случващото се с единствения главен участник, който може да бъде както одушевен, така и неодушевен. За антикаузативните глаголни единици са характерни три основни типа СЪБИТИЙНИ РАМКИ - ШАБЛОНИ, дадени съответно на Фигури 13, 14 и 15. В Таблицы 4, 5 и 6 са дадени част от включените в базата от данни антикаузативни глаголни единици на български език и техните екваленти на френски и унгарски език.

4.7. Дескриптори на пасивни композиции

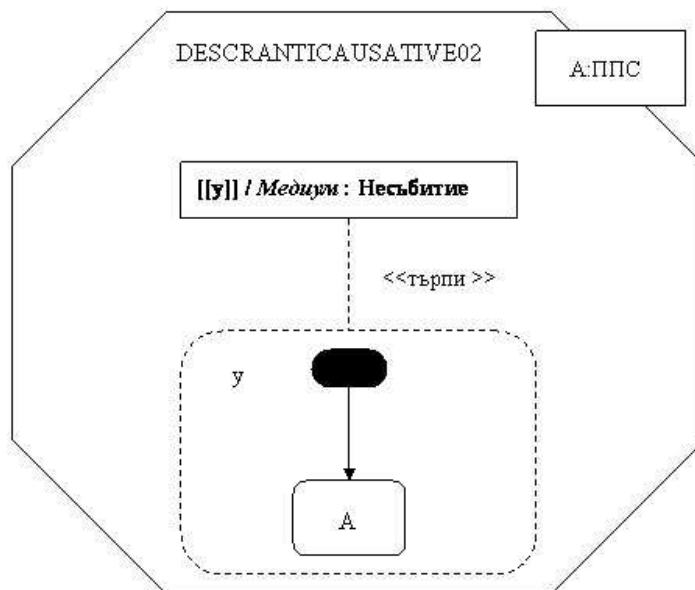
Пасивните глаголни композиции обозначават събития, в които акцентът е върху случващото се с неодушевен главен участник. Абстрактният *ролеви* КЛАС *Причинен* с подкласове Пациент и Тема е характерен за главния участник в пасивните композиции. На Фигура 16 е даден пример на СЪБИТИЙНА РАМКА - ШАБЛОН, която описва част от типичните представители на пасивни композиции като *Книгите се продават на улицата* – фр. *Les livres se vendent dans la rue; Креслото се поставя*



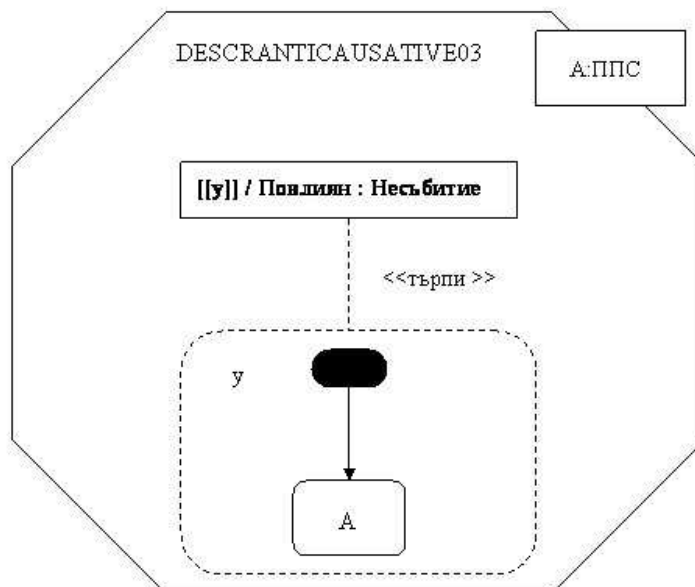
Фиг. 12. Дескриптор на абсолютни глаголни единици: вариант 1



Фиг. 13. Дескриптор на антикаузативни глаголни единици: вариант 1



Фиг. 14. Дескриптор на антикаузативни глаголни единици: вариант 2



Фиг. 15. Дескриптор на антикаузативни глаголни единици: вариант 3

АнтикаузБгИд	КорелатБг	АнтикаузБг	АнтикаузСъбРамШаблИд
ANTICAUSATIVEBG0001	вбесявам	вбесявам се	DESCRANTICAUSATIVE01
ANTICAUSATIVEBG0002	вдъхновявам	вдъхновявам се	DESCRANTICAUSATIVE01
ANTICAUSATIVEBG0003	влошавам	влошавам се	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0004	вцепенявам	вцепенявам се	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0005	възбуждам	възбуждам се	DESCRANTICAUSATIVE01
ANTICAUSATIVEBG0006	възмущавам	възмущавам се	DESCRANTICAUSATIVE01
ANTICAUSATIVEBG0007	възобновявам	възобновявам се	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0008	възраждам	възраждам се	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0009	възхищавам	възхищавам се	DESCRANTICAUSATIVE01
ANTICAUSATIVEBG0010	драскам	драскам се	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0011	драця	драця се	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0012	забавям	забавям се	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0013	заблуждавам	заблуждавам се	DESCRANTICAUSATIVE01
ANTICAUSATIVEBG0014	задръствам	задръствам се	DESCRANTICAUSATIVE03
ANTICAUSATIVEBG0015	задушавам	задушавам се	DESCRANTICAUSATIVE01
ANTICAUSATIVEBG0016	заледявам	заледявам се	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0017	залепям	залепям се	DESCRANTICAUSATIVE03
ANTICAUSATIVEBG0018	заоблям	заоблям се	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0019	запушвам	запушвам се	DESCRANTICAUSATIVE03
ANTICAUSATIVEBG0020	запълням	запълням се	DESCRANTICAUSATIVE03
ANTICAUSATIVEBG0021	заразявам	заразявам се	DESCRANTICAUSATIVE01
ANTICAUSATIVEBG0022	засягам	засягам се	DESCRANTICAUSATIVE01
ANTICAUSATIVEBG0023	затоплям	затоплям се	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0024	затруднявам	затруднявам се	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0025	избистрям	избистрям се	DESCRANTICAUSATIVE03
ANTICAUSATIVEBG0026	изгарям	изгарям се	DESCRANTICAUSATIVE01
ANTICAUSATIVEBG0027	изливам	изливам се	DESCRANTICAUSATIVE03
ANTICAUSATIVEBG0028	износвам	износвам се	DESCRANTICAUSATIVE03
ANTICAUSATIVEBG0029	изпичам	изпичам се	DESCRANTICAUSATIVE03
ANTICAUSATIVEBG0030	изпочупвам	изпочупвам се	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0031	изсушавам	изсушавам се	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0032	изтощавам	изтощавам се	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0033	изумявам	изумявам се	DESCRANTICAUSATIVE01
ANTICAUSATIVEBG0034	изхабявам	изхабявам се	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0035	изчерпвам	изчерпвам се	DESCRANTICAUSATIVE03
ANTICAUSATIVEBG0036	изчерпям	изчерпям се	DESCRANTICAUSATIVE03
ANTICAUSATIVEBG0037	лепвам	лепвам се	DESCRANTICAUSATIVE03
ANTICAUSATIVEBG0038	наводнявам	наводнявам се	DESCRANTICAUSATIVE03
ANTICAUSATIVEBG0039	напълвам	напълвам се	DESCRANTICAUSATIVE03
ANTICAUSATIVEBG0040	натрупвам	натрупвам се	DESCRANTICAUSATIVE03
ANTICAUSATIVEBG0041	обиждам	обиждам се	DESCRANTICAUSATIVE01
ANTICAUSATIVEBG0042	обезценявам	обезценявам се	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0043	обновявам	обновявам се	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0044	обучавам	обучавам се	DESCRANTICAUSATIVE01
ANTICAUSATIVEBG0045	огъвам	огъвам се	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0046	ожесточавам	ожесточавам се	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0047	оживявам	оживявам се	DESCRANTICAUSATIVE01
ANTICAUSATIVEBG0048	омотавам	омотавам се	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0049	опразвам	опразвам се	DESCRANTICAUSATIVE03
ANTICAUSATIVEBG0050	осведомявам	осведомявам се	DESCRANTICAUSATIVE01
ANTICAUSATIVEBG0051	освобождавам	освобождавам се	DESCRANTICAUSATIVE02

Таблица 4. Антикаузативни глаголни единици на български език: част 1

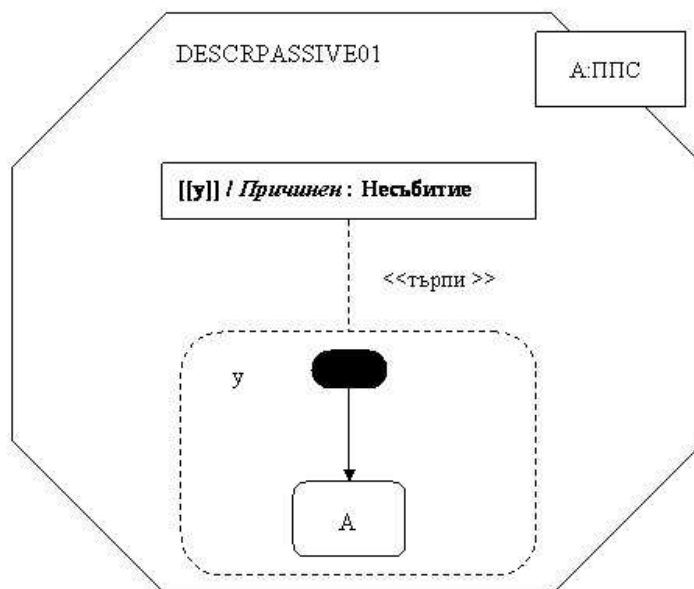
АнтикаузБгИд	КорелатФр01	АнтикаузФр01	АнтикаузСъбРамШаблИд
ANTICAUSATIVEBG0001	rendre furieux	devenir furieux	DESCRANTICAUSATIVE01
ANTICAUSATIVEBG0002	inspirer	s'inspirer	DESCRANTICAUSATIVE01
ANTICAUSATIVEBG0003	aggraver	s'aggraver	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0004	engourdir	s'engourdir	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0005	irriter	s'irriter	DESCRANTICAUSATIVE01
ANTICAUSATIVEBG0006	indigner	s'indigner	DESCRANTICAUSATIVE01
ANTICAUSATIVEBG0007	renouveler	se renouveler	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0008	faire renaître	renaître	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0009	enchanter	être enchanté	DESCRANTICAUSATIVE01
ANTICAUSATIVEBG0010	écorcher	s'écorcher	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0011	écorcher	s'écorcher	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0012	relentir	se relentir	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0013	égarer	s'égarer	DESCRANTICAUSATIVE01
ANTICAUSATIVEBG0014	engorger	s'engorger	DESCRANTICAUSATIVE03
ANTICAUSATIVEBG0015	étouffer	s'étouffer	DESCRANTICAUSATIVE01
ANTICAUSATIVEBG0016	geler	geler	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0017	coller	se coller	DESCRANTICAUSATIVE03
ANTICAUSATIVEBG0018	arrondir	s'arrondir	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0019	boucher	se boucher	DESCRANTICAUSATIVE03
ANTICAUSATIVEBG0020	remplir	se remplir	DESCRANTICAUSATIVE03
ANTICAUSATIVEBG0021	contaminer	attraper une maladie	DESCRANTICAUSATIVE01
ANTICAUSATIVEBG0022	offenser	s'offenser	DESCRANTICAUSATIVE01
ANTICAUSATIVEBG0023	réchauffer	se réchauffer	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0024	embarrasser	s'embarrasser	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0025	clarifier	se clarifier	DESCRANTICAUSATIVE03
ANTICAUSATIVEBG0026	brûler	se brûler	DESCRANTICAUSATIVE01
ANTICAUSATIVEBG0027	répandre	répandre	DESCRANTICAUSATIVE03
ANTICAUSATIVEBG0028	user	s'user	DESCRANTICAUSATIVE03
ANTICAUSATIVEBG0029	cuire	cuire	DESCRANTICAUSATIVE03
ANTICAUSATIVEBG0030	casser	se casser	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0031	sécher	sécher	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0032	affaiblir	s'affaiblir	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0033	étonner	s'étonner	DESCRANTICAUSATIVE01
ANTICAUSATIVEBG0034	user	s'user	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0035	épuiser	s'épuiser	DESCRANTICAUSATIVE03
ANTICAUSATIVEBG0036	épuiser	s'épuiser	DESCRANTICAUSATIVE03
ANTICAUSATIVEBG0037	coller	se coller	DESCRANTICAUSATIVE03
ANTICAUSATIVEBG0038	inonder	s'inonder	DESCRANTICAUSATIVE03
ANTICAUSATIVEBG0039	remplir	se remplir	DESCRANTICAUSATIVE03
ANTICAUSATIVEBG0040	accumuler	s'accumuler	DESCRANTICAUSATIVE03
ANTICAUSATIVEBG0041	offenser	s'offenser	DESCRANTICAUSATIVE01
ANTICAUSATIVEBG0042	dévaloriser	se dévaloriser	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0043	rafraîchir	se rafraîchir	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0044	instruire	s'instruire	DESCRANTICAUSATIVE01
ANTICAUSATIVEBG0045	courber	se courber	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0046	exaspérer	s'exaspérer	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0047	animer	s'animer	DESCRANTICAUSATIVE01
ANTICAUSATIVEBG0048	entortiller	s'entortiller	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0049	vider	se vider	DESCRANTICAUSATIVE03
ANTICAUSATIVEBG0050	informer	s'informer	DESCRANTICAUSATIVE01
ANTICAUSATIVEBG0051	libérer	se libérer	DESCRANTICAUSATIVE02

Таблица 5. Еквиваленти 01 на френски език на антикаузативни глаголни единици на български език: част 1

АнтикаузБгИд	КорелатУнг01	АнтикаузУнг01	АнтикаузСъбРамШаблИд
ANTICAUSATIVEBG0001	felbőszít	felbőszül	DESCRANTICAUSATIVE01
ANTICAUSATIVEBG0002	fellelkesít	fellelkesedik	DESCRANTICAUSATIVE01
ANTICAUSATIVEBG0003	rosszabbít	rosszabbodik	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0004	megmerevít	megmerevedik	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0005	felizgat	felizgatja magát	DESCRANTICAUSATIVE01
ANTICAUSATIVEBG0006	felháborít	felháborodik	DESCRANTICAUSATIVE01
ANTICAUSATIVEBG0007	megújít	megújul	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0008	feléleszt	feléled	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0009	elragadtat	el van ragadtatva	DESCRANTICAUSATIVE01
ANTICAUSATIVEBG0010	karmol	karmolja magát	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0011	karmol	karmolja magát	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0012	laszt	lassul	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0013	megtéveszt	eltéved	DESCRANTICAUSATIVE01
ANTICAUSATIVEBG0014	eldugaszol	betömődik	DESCRANTICAUSATIVE03
ANTICAUSATIVEBG0015	fullaszt	fulladozik	DESCRANTICAUSATIVE01
ANTICAUSATIVEBG0016	befagyaszt	befagy	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0017	ragaszt	ragad	DESCRANTICAUSATIVE03
ANTICAUSATIVEBG0018	gömbölyít	gömbölyödik	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0019	eltöm	eltömődik	DESCRANTICAUSATIVE03
ANTICAUSATIVEBG0020	betölt	betelik	DESCRANTICAUSATIVE03
ANTICAUSATIVEBG0021	megfertőz	megfertőződik	DESCRANTICAUSATIVE01
ANTICAUSATIVEBG0022	megbánt	megbántódik	DESCRANTICAUSATIVE01
ANTICAUSATIVEBG0023	melegít	melegszik	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0024	megnehezít	megnehezedik	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0025	tisztít	tisztul	DESCRANTICAUSATIVE03
ANTICAUSATIVEBG0026	éget	égeti magát	DESCRANTICAUSATIVE01
ANTICAUSATIVEBG0027	kiönt	kiömlik	DESCRANTICAUSATIVE03
ANTICAUSATIVEBG0028	koptat	kopik	DESCRANTICAUSATIVE03
ANTICAUSATIVEBG0029	megsüt	megsül	DESCRANTICAUSATIVE03
ANTICAUSATIVEBG0030	összetör	összetörik	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0031	kiszárít	kiszárad	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0032	kimerít	kimerül	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0033	elképeszt	elképed	DESCRANTICAUSATIVE01
ANTICAUSATIVEBG0034	elhasznál	elhasználódik	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0035	kimerít	kimerül	DESCRANTICAUSATIVE03
ANTICAUSATIVEBG0036	kimerít	kimerül	DESCRANTICAUSATIVE03
ANTICAUSATIVEBG0037	ragaszt	ragad	DESCRANTICAUSATIVE03
ANTICAUSATIVEBG0038	eláraszt	elárasztva van	DESCRANTICAUSATIVE03
ANTICAUSATIVEBG0039	megtölt	megtelik	DESCRANTICAUSATIVE03
ANTICAUSATIVEBG0040	felhalmoz	felhalmozódik	DESCRANTICAUSATIVE03
ANTICAUSATIVEBG0041	megsért	megsértődik	DESCRANTICAUSATIVE01
ANTICAUSATIVEBG0042	elértéktelenít	elértéktelenedik	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0043	feltűjt	feltűjul	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0044	oktat	tanul	DESCRANTICAUSATIVE01
ANTICAUSATIVEBG0045	meghajlít	meghajlik	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0046	elkeseredetté tesz	elkeseredetté válik	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0047	feléleszt	feléled	DESCRANTICAUSATIVE01
ANTICAUSATIVEBG0048	körülteker	körültekeredik	DESCRANTICAUSATIVE02
ANTICAUSATIVEBG0049	kiürít	kiürül	DESCRANTICAUSATIVE03
ANTICAUSATIVEBG0050	tájékoztat	tájékozódik	DESCRANTICAUSATIVE01
ANTICAUSATIVEBG0051	felszabadít	felszabadul	DESCRANTICAUSATIVE02

Таблица 6. Еквиваленти 01 на унгарски език на антикаузативни глаголни единици на български език: част 1

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 réaction s'observe parfois*.



Фиг. 16. Дескриптор 1 на пасивни композиции

4.8. Дескриптори на глаголни единици *корелат (-)*

За глаголите, наречени *reflexiva tantum*, е трудно да се намери обединяващ показател, който да ги вмести в единен семантичен клас. За разлика от останалите класове глаголи, представени в настоящата разработка, наличието им е случайно за всеки един от трите разглеждани езика и има формално значение за всеки един език поотделно, например, *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*. За семантичното описание на глаголите от този клас са използвани диаграми на СЪБИТИЙНИ РАМКИ - ШАБЛОНИ, които вече са представени в описанието на семантиката на глаголите от другите класове, затова идентификаторите на диаграмите съвпадат с тези от другите класове.

5. Заключение и перспективи

Развитието на информационните технологии води в настоящия момент до необходимостта, а оттам и до засиления интерес от създаването и развитието на модели и ресурси в областта на семантиката на езика. Освен това понастоящем се поставя

акцент върху разработването на многоезикови ресурси. Дисертационният труд е насочен към едни от най-актуалните теми в компютърната лингвистика и езиковите технологии, като при това представената в дисертационния труд разработка принадлежи към иновационното направление на *обектно-ориентираната семантика*. Създадените семантични дескриптори на *се/си*-глаголи в българския език и техни семантични еквиваленти във френския и унгарския език принадлежат към фундамента на езиковите технологии – моделирането на лингвистичното знание и изграждането на бази от знания за семантиката на езиковите обекти.

Поставените пред дисертационния труд цели и задачи са решени както следва.

Задача 1. *Дефиниране на класове глаголи в българския език, които са изходните единици в съпоставката с глаголи във френски и унгарски език.*

В изпълнение на тази задача е направен съпоставителен анализ на глаголни единици в българския, френския и унгарския език. Идентифицирани са *обектите на потребителя* като символни комбинации от форма и значение в българския език и на тях са съпоставени еквивалентни езикови обекти във френския и унгарския език. Глаголните единици първо се разделят в две групи: глаголни единици *корелат (+)* и глаголни единици *корелат (-)*. Глаголните единици *корелат (+)* се подразделят по-нататък в следните класове: глаголни единици със същинско рефлексивно значение; глаголни единици с реципрочно значение; глаголни единици, означаващи собствено движение; деакузативни глаголни единици; абсолютни глаголни единици; антикаузативни глаголни единици; възвратен пасив.

Задача 2. *Дефиниране на пакети от моделиращи елементи, необходими за описание на избраните езиковите обекти на концептуално равнище.*

На класовете обекти на потребителя са съпоставени модели на типове събития. Събитията са представени чрез диаграми, които включват статични моделиращи елементи, характеризиращи участниците в събитието и динамични моделиращи елементи, описващи поведението на участниците и взаимодействието между тях. Дефинирани са конфигурации от моделиращи елементи, които съответстват на концептуалните структури, кодирани в описаните в разработката класове глаголни единици.

Задача 3. *Разширяване на изходния модел с нови моделиращи елементи, необходими за семантичното описание на езиковите данни от настоящата разработка.*

Прилагането на модела върху представително множество от реални езикови данни наложи неговото разширяване. Въведени са нови моделиращи елементи, необходими за описанието както на статичния, така и на динамичния аспект на събитието. За попълното и адекватно представяне на КЛАСОВЕТЕ УЧАСТНИЦИ е дефинирана стройна система от РОЛИ НА УЧАСТНИЦИ, като е използвана възможността за определяне на семантични роли в йерархии. Въведен е специален абстрактен *ролеви* клас *Медиум* с подкласове *Експериенцер* и *Повлиян*. За моделирането на глаголни единици с реципрочно значение е въведено новото ОГРАНИЧЕНИЕ *Реципрочност*. Разширена е и изходната класификация на събития с нови типове взаимодействия между участници, както и с типове събития, свързани с поведението на един участник. Въведен е

и нов тип СИГНАЛ – СИГНАЛ-запазване.

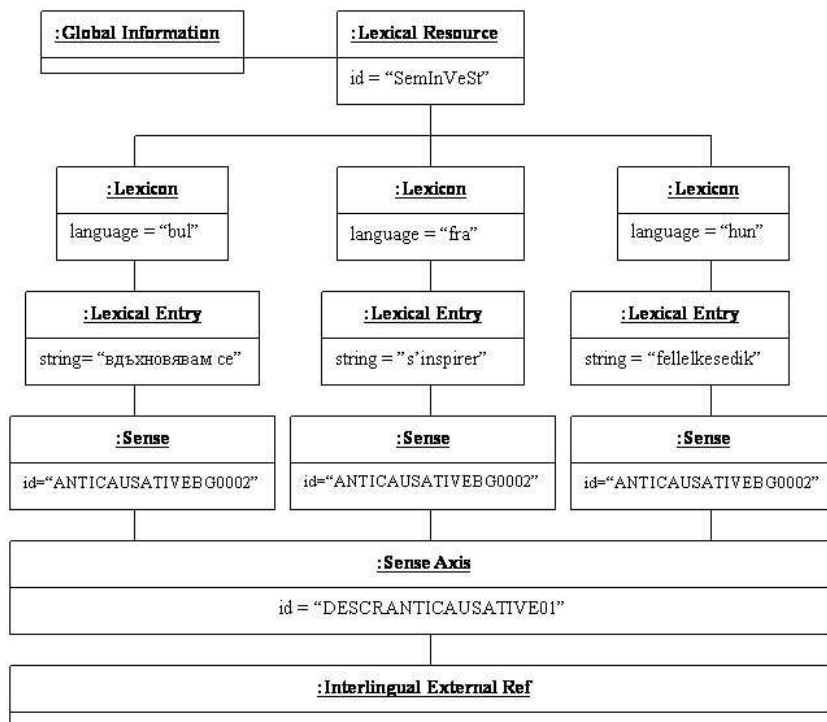
Задача 4. *Систематизиране на езиковите данни и построяване на диаграмите - семантични дескриптори.*

В дисертацията е представен продуктът **Семиinvest** – база от знания за семантиката на глаголни единици в българския език и техни еквиваленти във френския и унгарския език. Глаголните единици са подредени в таблици с данни, предназначени за свързване в една релационна база данни. Типът релация между данните на български и тези на френски или унгарски е *едно към много*, т.е. на един български глагол съответства един или повече от един глагол съответно на френски или на унгарски език, но на даден френски или унгарски глагол съответства точно един български глагол. Таблиците съдържат и полета на идентификаторите на СЪБИТИЙНИТЕ РАМКИ - ШАБЛОНИ, съответстващи на глаголите от таблицата. За всеки клас глаголни единици са съставени варианти на диаграми на СЪБИТИЙНИ РАМКИ - ШАБЛОНИ, които се асоциират с глаголите, посочени в съответния клас.

Предстояща задача е разширяването на базата от знания за глаголната семантика чрез включване на дескриптори на повече типове събития. Необходимо е да се подчертае, че представеният анализ и описание на рефлексивните по форма глаголни единици е свързан с описанието и на голям брой нереклексивни типове преходни и непреходни глаголи. Важен елемент на бъдещото разширение на **Семиinvest** е изграждането на по-голям брой специфицирани СЪБИТИЙНИ РАМКИ.

С оглед на включването на **Семиinvest** в глобални инфраструктури на езикови ресурси, се предвижда съвместяване (mapping) на **Семиinvest** с *Lexical Markup Framework* (LMF) – абстрактен метамодел, предоставящ стандартизирана среда за представяне на лексикални езикови ресурси (LMF 2008; Francopoulo et al. 2007). LMF е съгласуван със спецификациите и принципите на моделиране на UML. LMF се състои от *ядрен* моделиращ пакет (Core package) и моделиращи пакети с *разширения* (Extensions). За стандартизираното представяне на **Семиinvest** като многоезиков семантичен ресурс ще бъдат използвани *Семантичното* разширение и Разширението за *многоезикова нотация*. Например, на Фигура 17 е дадена диаграма в LMF формат на ниво екземпляр, която специфицира езикови обекти от **Семиinvest** чрез използване на разширението за многоезикова нотация. От практическа гледна точка е важна възможността за представяне на езиковия ресурс във формат на XML (eXtensible Markup Language).

Значимо бъдещо приложение на моделите на глаголната семантика, представени в дисертацията, е в областта на *извличането на информация* (Information Extraction (IE)) (Cunningham 2005) и по-специално *извличането на събития* (event extraction), известно също така като *извличане на сценарни шаблон* (scenario template extraction) (Liao and Grishman 2010; Grishman 2010). Целта на извличането на събития е идентифицирането в даден текст на екземпляри на даден клас събития. Идентифицирането на събитията е свързано с идентифицирането на неговите компоненти – участниците в събитието и различни негови атрибути. Извличането на събития предоставя средства за усъвършенстване на системите в областта на *отговарянето на въпроси*



Фиг. 17. LMF диаграма за многоезиков семантичен ресурс

(Question Answering) – поставят се по-сложни въпроси, чийто отговор изисква прилагането на методи, използващи по-абстрактни, структурирани, базирани на семантиката представяния на лингвистичното знание, за разлика от по-простите методи, основаващи се на търсенето по ключови думи (Grishman 2007; Schiffman et al. 2007). Многоезиковият аспект на разработката я прави приложима и в областта на *машинния превод* (Machine Translation).

ПРИНОСИТЕ на дисертацията са следните:

1. Предложена е цялостна инфраструктура за представяне на семантиката на езикови обекти в многоезиков план. Приложен е иновационен обектно-ориентиран подход при представяне на лингвистично знание с цел разширяване на възможностите за създаване на оптимални модели и подходящи езикови ресурси в областта на езиковите технологии.
2. Морфосинтактичното знание за глаголите е систематизирано и представено с оглед на интердисциплинарния характер на разработката и според принципите на изграждане на формални езикови модели и електронни езикови ресурси.
3. Направено е съпоставително изследване на значими в компютърната обработка типове глаголни единици – *се/cи*-глаголи в българския език и техните еквиваленти във френския и унгарския език.
4. Разширен е приложеният формализъм на ЕПС чрез добавяне на нови модели-

ращи елементи с цел адаптирането му за описание на семантиката на избраните реални езикови данни.

5. Създадена е база от знания за глаголната семантика в многоезиков план – *Семинвест*. Компонентите на *Семинвест* съставляват динамична система, която се изгражда според предпочитания понастоящем в областта на информационните технологии принцип на услуги по заявка на потребител, т.е. изграждане и предоставяне на компоненти съобразно конкретна и актуална задача.

Апробация на резултатите от дисертацията

По темата на дисертацията са публикувани петнадесет самостоятелни научни статии. Десет от тях са публикувани в специализирани международни издания, между които и в поредицата *Lecture Notes in Artificial Intelligence* на известното издателство за научна литература *Springer*. Резултатите, описани в тези статии, са представени многократно в доклади пред интернационална научна аудитория.

Доклади, изнесени на международни научни форуми:

1. International Conference on *Lexis and Grammar (LGC'10)*, Белград, Сърбия, септември 2010.
2. International Conference on *Text, Speech and Dialog (TSD'08)*, Бърно, Чехия, септември 2008.
3. International Conference on *Lexis and Grammar (LGC'08)*, l'Aquila, Италия, септември 2008.
4. International Conference on *Lexis and Grammar (LGC'07)*, Бонифачо, Корсика, Франция, октомври 2007.
5. International Conference on *Text, Speech and Dialog (TSD'06)*, Бърно, Чехия, септември 2006.
6. International Conference on *Language Resources and Evaluation (LREC'06)*, Геноа, Италия, май 2006.
7. International Conference on *Language Resources and Evaluation (LREC'04)*, Лисабон, Португалия, май 2004.
8. International Workshop on *Treebanks and Linguistic Theories (TLT'03)*, Växjö, Швеция, ноември 2003.
9. International Workshop on *Information Extraction for Slavonic and Other Central and Eastern European Languages*, в рамките на International Conference *Recent Advances in Natural Language Processing (RANLP'03)*, Боровец, България, септември 2003.
10. International Workshop on *Morphological Processing of Slavic Languages*, в рамките на *European Chapter of the Association for Computational Linguistics (EACL'03)*, Будапеща, Унгария, април 2003.
11. International Conference on *Corpus Linguistics (CL'03)*, Ланкастър, Великобритания, април 2003.
12. International Workshop on *Treebanks and Linguistic Theories (TLT'02)*, Созопол, България, септември 2002.

Доклади, изнесени на национални научни събития:

1. Осми национални славистични четения: *Славяните. Балканите. Европа*, Софийски университет „Св. Климент Охридски“, април 2008.
2. Седми национални славистични четения: *Славистика и общество*, Софийски университет „Св. Климент Охридски“, април 2004.
3. Шести национални славистични четения: *Славистиката в началото на XXI век. Традиции и очаквания*, Софийски университет „Св. Климент Охридски“, април 2002.

ПУБЛИКАЦИИ ПО ДИСЕРТАЦИЯТА

Slavcheva, Milena. 2010. Verb Predicates: Concepts, Model Elements and Lexicon Structures. In *Actes du "27e Colloque international sur le lexique et la grammaire"*, Seconde partie, De Gioia, Michele (ed.), Collection : *Lingue d'Europa e del Mediterraneo. Grammatica comparata*, 1, Roma: Aracne, pp. 111-125. ISBN 978-88-548-3166-7

Slavcheva, Milena. 2010. Eventivity Type Taxonomy for a Lexical Semantic Knowledge Base. In *Proceedings of the 29th International Conference on Lexis and Grammar*, Vitas, Dusko and Krstev, Cvetana (eds.), Faculty of Mathematics, University of Belgrade, Belgrade, pp.311-320. ISBN 978-86-7589-080-5

Slavcheva, Milena. 2008. Thinking in Objects: Towards an Infrastructure for Semantic Representation of Verb-Centered Structures. In *TSD 2008, Lecture Notes in Artificial Intelligence, 5246*, Sojka, Petr, Horak, Ales, Kopecek, Ivan and Pala, Karel (eds.), Berlin/Heidelberg: Springer, pp. 193-200. ISSN 0302-9743; ISBN-10: 3-540-87390-2 Springer Berlin Heidelberg New York; ISBN-13: 978-3-540-87390-7 Springer Berlin Heidelberg New York.

Slavcheva, Milena. 2007. Linking Reflexive Verb Structure to Verb Meaning in a Cross-Lingual Lexical Setting. In *Proceedings of the 26th International Conference on Lexis and Grammar*, Bonifacio, Corse, France, Camugli, Catherine, Constant, Matthieu and Dister, Anne (eds.), Universite de Marne-la-Vallee, Institut Gaspard-Monge, pp.153-160.

Slavcheva, Milena. 2006. Semantic Representation of Events: Building a Semantic Primes Component. In *TSD 2006, Lecture Notes in Artificial Intelligence, 4188*, Sojka, Petr, Kopecek, Ivan and Pala, Karel (eds.), Berlin/ Heidelberg: Springer, pp. 245-252. ISSN 0302-9743; ISBN-10: 3-540-39090-1 Springer Berlin Heidelberg New York; ISBN-13: 978-3-540-39090-9 Springer Berlin Heidelberg New York.

Slavcheva, Milena. 2006. Semantic Descriptors: The Case of Reflexive Verbs. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC 2006)*, Genoa, Italy, pp. 1009-1014. Copyright by the European Language Resources Association. ISBN 2-9517408-2-4; EAN 9782951740822

Славчева, Милена. 2006. Възвратните по форма конструкции в българския език с оглед на автоматичния им анализ. *Славистика и общество*, М. Младенова, И. Монова (съставители), Херон Прес, София, с. 261-267. ISBN-10: 954-580-189-1; ISBN-13: 978-954-580-189-1

Slavcheva, Milena. 2006. Building Cross-Lingual Semantic Components: The Case of Reflexives. In *Readings in Multilinguality*, Slavcheva, Milena, Angelova, Galia and Simov,

Kiril (eds.), Incoma Ltd., Shoumen, pp.117-124. ISBN-10: 954-91743-6-0, ISBN-13: 978-954-91743-6-6

Slavcheva, Milena. 2004. Verb Valency Descriptors for a Syntactic Treebank. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC 2004)*, Lisboa, Portugal, pp. 1153-1156. Copyright by the European Language Resources Association, ISBN 2-9517408-1-6; EAN 9782951740815; 211 156/04

Slavcheva, Milena. 2003. Integrating a Verb Lexicon into a Syntactic Treebank Production. In *Proceedings of the Second Workshop on Treebanks and Linguistic Theories*, Nivre, Joakim and Hinrichs, Erhard (eds.), Series: *Mathematical Modelling in Physics, Engineering and Cognitive Sciences*, v.9, Vaxjo University Press, Sweden, pp.165-176. ISBN 91-7636-394-5; ISSN 1651-0267

Slavcheva, Milena. 2003. Extracting Verb Complex Structures in Bulgarian. In *Proceedings of the International Workshop on Information Extraction for Slavonic and Other Central and Eastern European Languages, RANLP 2003*, Cunningham, Hamish, Paskaleva, Elena, Bontcheva, Kalina and Angelova, Galia (eds.) Borovets, Bulgaria, pp. 94-101. ISBN 954-8361-06-X

Slavcheva, Milena. 2003. Some Aspects of the Morphological Processing of Bulgarian. In *Proceedings of the Workshop on Morphological Processing of Slavic Languages, EAACL 2003*, Budapest, Hungary, pp.71-77. Copyright by the Association for Computational Linguistics.

Slavcheva, Milena. 2003. Corpus Shallow Parsing: Meeting Point Between Paradigmatic Knowledge Encoding and Syntagmatic Pattern Matching. In *Proceedings of the Corpus Linguistics 2003 Conference*, Archer, Dawn, Rayson, Paul, Andrew Wilson, Andrew and McEnery, Tony (eds.), University Centre for Computer Corpus Research on Language, Lancaster University, Technical Papers Volume 16 - Special issue, Lancaster, UK, pp.706-713. ISBN 1-86220-131-5

Славчева, Милена. 2003. Езиковите технологии и българският език - един класификационен модел на глагола. В *Славистиката в началото на XXI век. Традиции и очаквания*, Издателска къща СемаРШ, София, с. 209-216. ISBN 954-8021-13-7

Slavcheva, Milena. 2002. Segmentation Layers in the Group of the Predicate: a Case Study of Bulgarian within the BulTreeBank Framework. In *Proceedings of the International Workshop on Treebanks and Linguistic Theories*, Sozopol, Bulgaria, pp.199-210.

Дисертационният труд е компонент от цялостната ми научно-изследователска и научно - приложна работа, видима от участието ми в **ПРОЕКТИ** като:

AsIsKnown (2006-2009): Система за базирано върху семантиката движение на знание за европейската текстилна индустрия (A semantic-based knowledge flow system for the European home textiles industry), финансиран от Европейската комисия в Шестата рамкова програма, Специфична програма: IST-2004-2.4.7 Тематичен приоритет / Хоризонтална дейност: Semantic-based Knowledge and Content Systems.

Моята задача в рамките на проекта бе създаване на терминологичен тълковен лекси-

кон на унгарски език – компонент от ресурсното осигуряване на система в областта на домашния текстил и свързаните с него области архитектура и вътрешен дизайн. Лексиконът съдържа 2500 терминологични единици, снабдени с дефиниции на унгарски език.

BIS-21++ (2005-2007): Център за компетентност в 21 век (Centre of Competence in 21 Century), финансиран от Европейската комисия в Шестата рамкова програма (FP6-2004-ACC-SSA-2), Работен пакет *Езикови технологии*

BulTreeBank (2001-2004): Банка от синтактични дървовидни структури на българския език, базирана върху Опорно- фразовата граматика (HPSG-based Syntactic Treebank of Bulgarian), финансиран от Фондация Volkswagen, Германия.

Моята задача в рамките на проекта бе създаване и прилагане на частичен синтактичен автоматичен анализ, който идентифицира в текста сегменти, съставени от пълнозначен глагол, спомагателни глаголи и клитики.

BIS-21 (2001-2004): Център с отлични постижения в 21 век (Centre of Excellence in 21 Century), финансиран от Европейската комисия в Петата рамкова програма, Работен пакет 5: *Инструменти за лингвистични изследвания, базирани върху знания.*

Проект на гост-учен в *Equipe langue et dialogue, LORIA*, Нанси, Франция (2003): *Категории данни за синтактична лингвистична анотация.*

Проектът е свързан с глобалния проблем за съвместимост и преносимост на кодирана лингвистична информация. Извършени са експерименти за интегриране на подмножество на специфични категории за синтактично аотиране в *Регистъра на категории от данни* (Data Category Registry) в рамките на модела за обобщени схеми за кодиране на лингвистична информация, разработен в *INRIA Lorraine*, Франция. Изследвана е приложимостта на модела за анотационни типове върху конкретни синтактични структури чрез тяхното описание на няколко нива на абстракция. За целта са описани типове предикатно-аргументни структури, извлечени от изречения на естествен език. Съставени са мета- граматика като дескриптивен модул за включване на специфичните категории от данни към обобщения модел за кодиране на лингвистична информация.

TELRI II (1999-2002): Трансевропейска инфраструктура за езикови ресурси (Trans-European Language Resources Infrastructure), финансиран от Европейската комисия, Програма: COPERNICUS'94 Concerted Action 1202.

Една част от задачите ми по този проект (свързани с Работна група 3 *Морфосинтактично аотиране*) бяха: създаване на български текстов корпус в стандартизиран формат (Corpus Encoding Standard (CES)) за *TRACTOR: TELRI Research Archive of Computational Tools and Resources* (Архив на *TELRI* за компютърни инструменти и ресурси); сравнително изследване на схеми за кодиране на морфосинтактична информация; приложение за българския език на стандарта *EAGLES* за формализирано представяне на морфосинтактични явления в лексикони и текстови корпуси;

създаване на модели на текстове от различни жанрове и дефиниране на стандартизирани анотационни схеми за изграждането на *TENCO* (Automatic Text Encoder) – система за автоматично анотиране в XML формат на текстове с цел изграждане на електронни текстови ресурси (съвместна разработка с Боянка Захариева).

Друга част от задачите ми бяха свързани с работата ми като гост-учен и мениджър на проекта при Координатора на проекта в *Института за немски език* (Institut für deutsche Sprache (IDS)) в Манхайм, Германия.

ELAN (1999-2000): Европейска мрежа за лингвистична дейност (European Language Activity Network), финансиран от Европейската комисия.

Работата по проекта бе свързана с валидиране на стандартизирани текстови и лексикални ресурси от 21 академични институции, разположени в страните на Централна и Източна Европа. Работното ми място по този проект бе в *Института за немски език* (Institut für deutsche Sprache (IDS)) в Манхайм, Германия.

GeFRePaC (1999-2000): Немско-френски реципрочен паралелен корпус (German - French Reciprocal Parallel Corpus).

Работата по проекта бе свързана с изграждане на корпус от немски и френски паралелни текстове, които са морфосинтактично анотирани, подравнени на ниво изречение и анотирани съгласно схемите на TEI (Text Encoding Initiative) за стандартизирано представяне на паралелни текстове. Работното ми място по този проект бе в *Института за немски език* (Institut für deutsche Sprache (IDS)) в Манхайм, Германия.

BILEDITA (1995-1997): Двуетични електронни речници и интелигентно подравняване на текстове (Bilingual Electronic Dictionaries and Intelligent Text Alignment), финансиран от Европейската комисия, Програма: COPERNICUS'94 Joint Research Project 790.

GLOSSER (1995-1997): Подпомагане на усвояването на втори език и усвояването на знания от подравнени паралелни корпуси (Support of Second Language Acquisition and Learning from Aligned Corpora), финансиран от Европейската комисия, Програма: Copernicus'94 Joint Research Project 343.

Голяма лексикална база от данни за българския език (1989-1999), научен ръководител доц. д-р Елена Паскалева.

По този проект съм участвала в изграждането на формален морфологичен модел на българския език и съставяне на база от лексеми и техните словоформи, всяка снабдена с морфосинтактична информация. Акцентът на работата ми бе върху класификационен модел на глаголите в българския език, основаващ се на морфосинтактичните признаци, характеризиращи глаголните лексеми.

Литература

Бондарко, А. В. (ред.). *Теория функциональной грамматики. Персональность. Залоговость*. Издательство „Наука“, Санкт-Петербург, 1991.

Бондарко, А. В. (ред.). *Проблемы функциональной грамматики: Полевые структуры*. Издательство „Наука“, Санкт-Петербург, 2005.

Ботева, Силвия. *Глаголът във френския и българския език. Функционално-семантична граматика*. Издателска къща КОЛИБРИ, София, 2000.

Паскалева, Елена. *Компютърна морфология. Ресурси и инструменти*. Институт за паралелна обработка на информацията, БАН, София, 2007.

Славчева, Милена. Езиковите технологии и българският език — един класификационен модел на глагола. В *Славистиката в началото на XXI век. Традиции и очаквания*, стр. 209—216. Издателска къща СемаРШ, София, 2003.

Фаулър, Мартин. *UML. Основи*. Издателство СофтПрес, София, 2004.

Angelova, Galia. Ontologies for natural language processing applications. In Simov, Kiril and Kiryakov, Atanas, editors, *Ontologies and Lexical Knowledge Bases. Proceedings of OntoLex 2000*, pages 1–15. 2000.

Angelova, Galia. Language technologies meet ontology acquisition. In Mugnier M.-L. Dau, F. and G. Stumme, editors, *Conceptual Structures: Common Semantics for Sharing Knowledge. Proceedings of ICCS 2005, Lecture Notes in Artificial Intelligence 3596*, pages 367–380. Springer, Sozopol, Bulgaria, 2005.

Dobrev, Pavlin, Boytcheva, Svetla and Angelova, Galia. Cgextract: Towards extraction of conceptual graphs from controlled english. In G. Mineau, editor, *Conceptual Structures: Extracting and Representing Semantics. Proceedings of ICCS 2001*, pages 89–116. CEUR Workshop Proceedings, Stanford, California, 2001.

Cunningham, Hamish. Information Extraction, Automatic. In *Encyclopedia of Language and Linguistics, 2nd Edition*. Elsevier, 2005.

Dowty, David R. Thematic proto-roles and argument selection. *Language*, 67(3): 547–619, 1991.

Evans, Vyvyan and Green, Melanie. *Cognitive Linguistics: an Introduction*. Edinburgh University Press, UK, 2006.

Francopoulo, Gil, Bel, Nuria, George, Monte, Calzolari, Nicoletta, Monachini, Monica, Pet, Mandi and Soria, Claudia. Lexical markup framework: ISO standard for semantic information in NLP lexicons. In *(Gesellschaft fuer linguistische Datenverarbeitung) (GLDV)*. Tuebingen, Germany, 2007.

Grishman, Ralph. Extracting information and answering questions. In *Proceedings of the First International Symposium on Universal Communication*. Kyoto, Japan, 2007.

Grishman, Ralph. The impact of task and corpus on event extraction systems. In *Proceedings of the Seventh Conference on Language Resources and Evaluation (LREC 2010)*, pages 2928–2931. European Language Resources Association (ELRA), Valletta, Malta, 2010.

Guentchéva, Zlatka and Rivière, Nicole. Reciprocal and reflexive constructions in French. In Nedjalkov, Vladimir P., editor, *Reciprocal Constructions*, volume 2 of *Typological Studies in Language 71*, chapter 12, pages 561–608. John Benjamins Publishing Company, 2007.

Kiryakov, Atanas. Ontologies for knowledge management. In *Semantic Web Technologies: Trends and Research in Ontology-based Systems*, page 115. 2006.

Lakoff, George. Cognitive models and prototype theory. In U. Neisser, editor, *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization*. Cambridge University Press, 1987.

Liao, Shasha and Grishman, Ralph. Using document level cross-event inference to improve event extraction. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 789–797. Association for Computational Linguistics, Uppsala, Sweden, 2010.

LMF. *ISO 24613: Language resource management - Lexical markup framework (LMF)*. 2008.

OMG. *OMG Unified Modeling Language Specification, Version 1.4*. Object Management Group (OMG), <http://www.omg.org/>, 2001.

Paskaleva, Elena. Compilation and validation of morphological resources (overview of the morphology cooking technologies). In Piperidis, Stelios and Karkaletsis, Vangelis, 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.

Paskaleva, Elena, Simov, Kiril, Damova, Mariana and Slavcheva, Milena. The long journey from the core to the real size of large LDBs. In Boguraev, Branimir and Pustezovsky, James, 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.

Penchev, Jordan. Reciprocal and reflexive constructions in Bulgarian. In Nedjalkov, Vladimir P., editor, *Reciprocal Constructions*, volume 2 of *Typological Studies in Language 71*, chapter 13, pages 609–632. John Benjamins Publishing Company, 2007.

- Popov, B., Kiryakov, A., Kirilov, A., Manov, D., Ognyanoff, D. and Goranov, M. Kim - semantic annotation platform. In *Second International Semantic Web Conference (ISWC2003)*, LNAI 2870, pages 834–849. Springer-Verlag, 2003.
- Schalley, Andrea C. *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.
- Schiffman, Barry, McKeown, Kathleen R., Grishman, Ralph and Allan, James. Question answering using integrated information retrieval and information extraction. In *Proceedings of NAACL HLT 2007*, pages 532–539. Association for Computational Linguistics, Rochester, NY, 2007.
- Simov, Kiril. Ontology-based lexicon of bulgarian. *LDV Forum*, 21(1):1–18, 2006.
- Simov, Kiril and Osenova, Petya. Resources and tools for ontology-based semantic annotation. In *Proceedings of Workshop OntoLex 2008, LREC 2008*, pages 9–13. Marrakech, Morocco, 2008.
- Simov, Kiril, Osenova, Petya, Slavcheva, Milena, Kolkovska, Sia, Balabanova, Elisaveta, Doikov, Dimitar, Ivanova, Krasimira, Simov, Alexander and Kouylekov, Milen. 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.
- Slavcheva, Milena. Segmentation layers in the group of the predicate: a case study of Bulgarian within the BulTreeBank framework. In Hinrichs, Erhard and Simov, Kiril, editors, *Proceedings of the First Workshop on Treebanks and Linguistic Theories (TLT 2002)*, pages 199–210. Sozopol, Bulgaria, 2002.
- Slavcheva, Milena. 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.
- Slavcheva, Milena. Corpus shallow parsing: Meeting point between paradigmatic knowledge encoding and syntagmatic pattern matching. In *Proceedings of the Corpus Linguistics Conference 2003*. Lancaster, UK, 2003b.
- Slavcheva, Milena. Extracting verb complex structures in Bulgarian. In Cunningham, Hamish, Paskaleva, Elena, Angelova, Galia and 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.
- Talmy, Leonard. *Toward a Cognitive Semantics. Vol. 2: Typology and Process in Concept Structuring*. MIT Press, Cambridge, MA / London, 2000.

Abstracts of Dissertations

Number 1, 2012

INSTITUTE OF INFORMATION AND COMMUNICATION TECHNOLOGIES
BULGARIAN ACADEMY OF SCIENCES

БЪЛГАРСКА АКАДЕМИЯ НА НАУКИТЕ
ИНСТИТУТ ПО ИНФОРМАЦИОННИ И КОМУНИКАЦИОННИ ТЕХНОЛОГИИ

Брой 1, 2012

Автореферати на дисертации