

Semantic Descriptors: The Case of Reflexive Verbs

Milena Slavcheva

Bulgarian Academy of Sciences
Institute for Parallel Processing
25A, Acad. G. Bonchev St., 1113 Sofia, Bulgaria
milena@lml.bas.bg

Abstract

This paper presents a semantic classification of reflexive verbs in Bulgarian, augmenting the morphosyntactic classes of verbs in the large Bulgarian Lexical Data Base - a language resource utilized in a number of Language Engineering (LE) applications. The semantic descriptors conform to the Unified Eventivity Representation (UER), developed by Andrea Schalley. The UER is a graphical formalism, introducing the object-oriented system design to linguistic semantics. Reflexive/non-reflexive verb pairs are analyzed where the non-reflexive member of the opposition, a two-place predicate, is considered the initial linguistic entity from which the reflexive correlate is derived. The reflexive verbs are distributed into initial syntactic-semantic classes which serve as the basis for defining the relevant semantic descriptors in the form of EVENTIVITY FRAME diagrams. The factors that influence the categorization of the reflexives are the lexical paradigmatic approach to the data, the choice of only one reading for each verb, top level generalization of the semantic descriptors. The language models described in this paper provide the possibility for building linguistic components utilizable in knowledge-driven systems.

1. Introduction

This paper presents a semantic classification of reflexive verbs in Bulgarian, augmenting the morphosyntactic classes of verbs (Slavcheva, 2004) in the large Bulgarian Lexical Data Base - a language resource utilized in a number of Language Engineering (LE) applications (Paskaleva et al., 1993; Paskaleva, 2003a; Slavcheva, 2003a; Slavcheva, 2003b). The semantic descriptors conform to the Unified Eventivity Representation (UER), developed by Andrea Schalley (Schalley, 2004). The UER is a graphical formalism, introducing the object-oriented system design to linguistic semantics. The UER is based on the Unified Modeling Language (UML) (OMG) - "the current lingua franca for the design of object-oriented systems in computer science" (Schalley, 2004).

The task of supplying verbs with elaborate semantic descriptions is rather ambitious and it should be subdivided into realistically defined subtasks. Although the UER framework deals with the "deep" semantics of verbs and starts from the purely semantic level of analysis, in a real-world application as the current one, it is necessary to bring the semantic knowledge to anchors, which, being formal, are easier to define. Such an anchor for Bulgarian are the pairs of reflexive verbs and their non-reflexive counterparts, where "reflexive" is used to name the polysemantic construct including a verb and a reflexive pronominal clitic. The investigation of the morphosyntax and semantics of reflexives is relevant to a great number of typologically different languages (Genyushene, Nedyalkov, 1991).

The reflexive verb forms acquire a variety of functions ranging from grammatical formation (e.g., the morphological expression of verbal voice like *passive*) to lexical derivation and that is why they are difficult to represent systematically in a large-scale application. At the same time, their systematic semantic interpretation is a considerable portion of the semantic knowledge representation of a given language. If we imagine the reflexive verb forms distributed on a scale depending on the degree of their lexicalization,

the current research deals with those reflexives which coincide with, or approximate the lexicalized extremes of the scale.

The paper is structured as follows. Section 2 represents an initial syntactic-semantic classification of reflexive verbs derived from transitive non-reflexive verbs. In Section 3 semantic descriptors in the form of diagrams are provided for the currently defined semantic classes of reflexive verbs in Bulgarian. Section 4 contains some results and a brief discussion, as well as clues for further development.

2. Initial classification

Reflexive/non-reflexive verb pairs are analyzed where the non-reflexive member of the opposition, a two-place predicate, is considered the initial linguistic entity from which the reflexive correlate is derived. The reflexive verbs are distributed into initial syntactic-semantic classes based on those of (Genyushene, Nedyalkov, 1991), regarding the following factors. The primary discriminating feature of reflexives is the *subject-centrality*, "*subject-drivenness*" of the action denoted by the verb, and it is relevant to observe which argument in the non-reflexive structure is promoted to the subject position in the reflexive structure, or is affected in some way so that some sort of a reflexive structure is produced (Genyushene, Nedyalkov, 1991; Kordi, 1981; Boteva, 2000). This transition is concurrent with the elimination or change of the status of another argument or other arguments of the non-reflexive. Thus the two *top-level* classes of the hierarchy, relevant for the current investigation, are: *subject-retaining* reflexives (the subject position argument of the non-reflexive preserves its subject position in the reflexive); *object-derived* reflexives (the object position argument of the non-reflexive becomes a subject position argument in the reflexive). Taking into account some semantic features of the participants and the action, like *animacy*, *volition*, *spontaneity*, *instigation*, *sentience*, *causation*, the verbs are further distributed into classes which serve as the basis for defining the relevant EVENTIVITY

FRAME diagrams. The top-level class of subject-retaining reflexives is subdivided into *inherent* (or semantic), *motive*, *absolutive*, *deaccusative* reflexives. The class of object-derived reflexives consists of *decausatives*.

3. Semantic descriptors

The UER is a framework which tries to achieve cognitive adequacy of the representation of verbal semantics and as such "understands the meaning of verbs to be conceivable as concepts of events and similar entities in the mind." (Schalley, 2004, p.1) Thus the central concept of the framework is defined as an *eventity* and is represented by an EVENTITY FRAME diagram, which contains a *dynamic core* and a *static periphery*. The dynamic core is a state chart depicting the state-transition system of the conceptualized actions. The static periphery includes representation of the participants, their properties and relations. The participants' specifications refer to PARTICIPANT CLASSES whose properties are described in sets of ATTRIBUTES. (Schalley, 2004)

The modeling elements, contained in the EVENTITY FRAME diagram, which determine the parameters according to which the verb predicates are semantically classified and supplied with an appropriate descriptor, are:

- PARTICIPANT CLASSES and PARTICIPANT OBJECTS which depict the number of prominent participants and non-prominent participants involved;
- PARTICIPANT ROLES, which reference the semantic roles of participants;
- PARTICIPANT TYPES, which reference the ontological type of participants in a hierarchy;
- ASSOCIATION relation, which models the relations between participants in the static periphery;
- PARTICIPATE ASSOCIATION, which models the relation between a participant and the dynamic core;
- state-machines, which depict the character of the action and the interaction between participants in the dynamic core;
- TEMPLATES, which are parameterized model elements: contain slots to be filled with values and model systematic relations among linguistic entities.

At present semantic description is provided for a data set of Bulgarian verbs consisting of 337 lexicon units selected via their morphosyntactic features from a Bulgarian lexical data base of 10,216 lemmas (an expanded subset of the large Bulgarian Lexical Data Base (60,000 lemmas) produced as the result of the processing of frequency lists of wordforms occurring in newspaper texts (5 million wordforms)(Paskaleva, 2003b). The morphosyntactic features of the lexicon units are: *verb, personal, perfective, transitive, possible attachment of "se"* (*se* is the clitic form of the reflexive accusative pronoun). The particularly selected data set is representative of the investigated language phenomenon.

In the following subsections the currently defined semantic classes of reflexives in Bulgarian are presented.

3.1. Inherent reflexives

Figure 1 provides a generic EVENTITY FRAME diagram (an octagon container) of the prototypical initial non-reflexive verb predicate from which an inherent reflexive is derived. The rectangles in the upper part of the octagon belong to the static periphery of the EVENTITY FRAME and provide information for the two prominent participants, whose semantic roles are specified as Agent and Patient respectively, and whose ontological categories (the PARTICIPANT TYPES) are Individual and Ineventity respectively. The PARTICIPANT TYPES are referenced to the present UER participant ontology (Schalley, 2004, p.197). 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) depicts the state-machines and the causation interaction (via the cause-SIGNAL) between the prominent participants, each one of which has its own SWIMLANE (indicated by the solid vertical line separating the dashed-outline rectangle with rounded corners). The left SWIMLANE belongs to the Agent, and its state-machine is interpreted as follows. At some point in time the Agent, being in an unspecified Active Simple State (ASS) sends a cause-SIGNAL, notated by the solid arrow crossing the borderline and entering the right SWIMLANE belonging to the Patient. The cause-SIGNAL triggers a transition of the second participant from an unspecified source state to a parameterized target state which in this case is generalized as being a Passive Simple State (PSS). The dash-outline rectangle in the upper right corner of the octagon indicates that the EVENTITY FRAME is a TEMPLATE which has a parameter to be bound.

The interpretation of the generic EVENTITY FRAME of the prototypical initial non-reflexive verb predicate from which an inherent reflexive is derived is "an action of an animate agent is directed towards an inanimate or animate patient, the relation between the two prominent participants is that of causation starting from the agent and directed towards the patient".

Figure 2 is the semantic model of the class of the inherent reflexives. It is interpreted as "an action of an animate agent is directed towards himself; the nature of the causation relation is preserved, but the cause-signal starts from an agent participant and returns back to him." It should be noted that the quantity of inherent reflexives as part of the lexicon fluctuates as the issue of truly lexicalized predicates and occasional reflexives always comes into play.

Examples of inherent reflexives are: *obleka se* ('dress oneself'), *izmija se* ('wash oneself'), *spasja se* ('save oneself'), *hvalja se* ('boast').

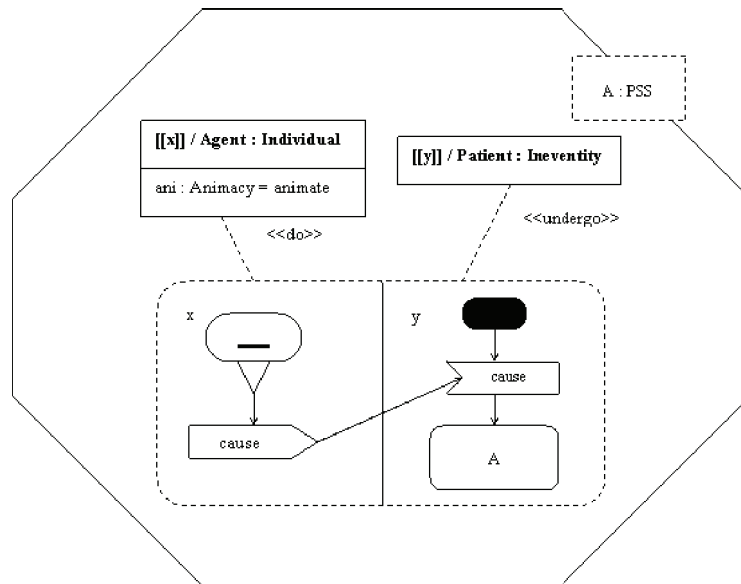


Figure 1: Transitive EVENTITY FRAME.

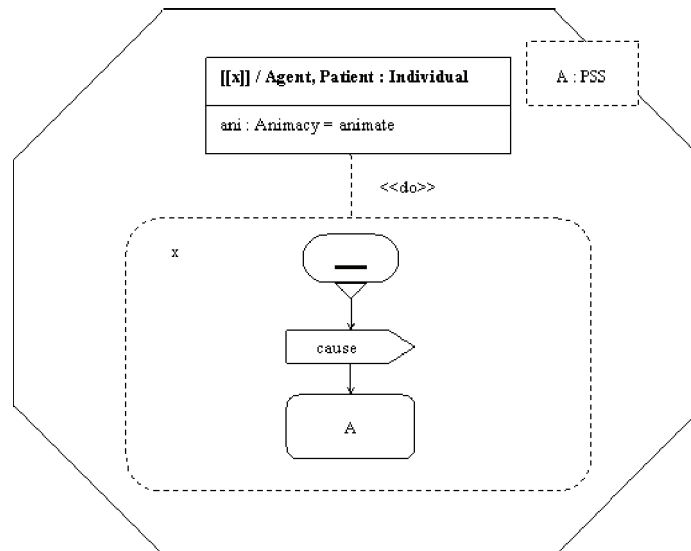


Figure 2: Inherent reflexives.

3.2. Motive reflexives

Figure 3 provides the prototypical EVENTITY FRAME of the motive reflexives. The single participant is generically defined as the actor who's body is intrinsically and fully involved in the action. The state machine in the dynamic core models a transition that is conceptualized as inner to the participant, hence the absence of a cause-SIGNAL.

Examples of motive reflexives are: *skrija se* ('hide oneself'), *kača se* ('go up, climb'), *nastanja se* ('settle oneself').

3.3. Absolute reflexives

Figure 4 represents the EVENTITY FRAME diagram of the absolute reflexives. They are predominantly related to mental activities, activities of the will, social activities, etc. In the dynamic core the mentality sense is depicted by

a transition from an unspecified source state. Prototypically the meaning of the absolute reflexives is analogous to the meaning of absolutely used transitive verbs like *eat*, *read*, etc. (*He eats a sandwich. / He eats.*).

Examples of absolute reflexives are: *izkaža se* ('express oneself'), *nasitja se* ('be full, be sated'), *otârva se* ('rid oneself').

3.4. Deaccusative reflexives

In the case of deaccusative reflexives the status of the undergoer changes from that of a prominent participant in the eventy describing the initial, transitive, non-reflexive predicate to that of a non-prominent, but conceptualized participant in the eventy describing the derived reflexive predicate. Figure 5 provides the prototypical EVENTITY FRAME of the deaccusative reflexives.

Examples of deaccusatives are: *približa se* ('get nearer to),

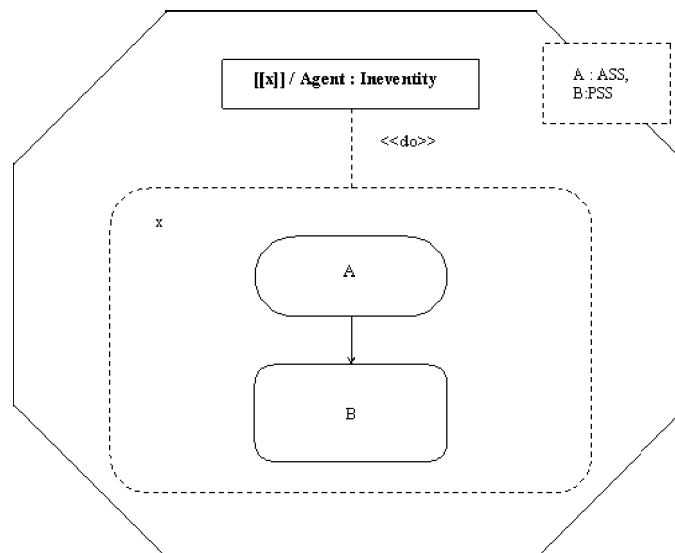


Figure 3: Motive reflexives.

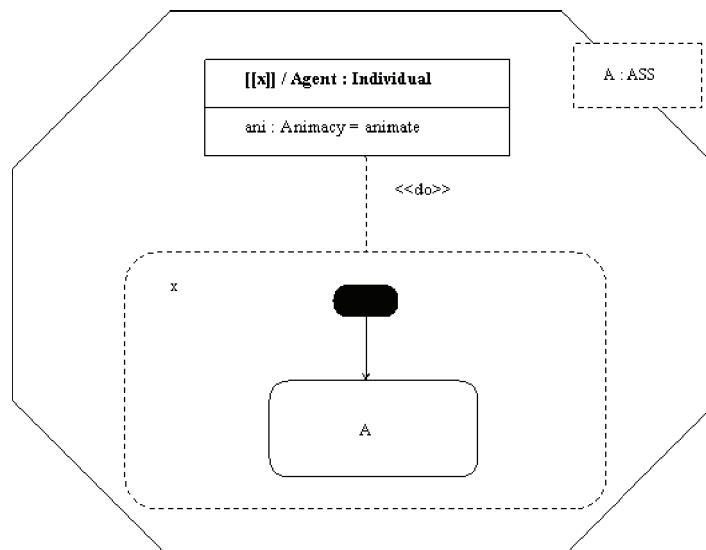


Figure 4: Absolute reflexives.

pomolja se ('ask for').

3.5. Decausative reflexives

The decausative reflexives are derived from initial transitive non-reflexive predicates, where the actor PARTICIPANT ROLE is Instigator - a generalized semantic role that is the parent of the semantic roles Agent (volitional) and Effector (non-volitional) (Schalley, 2004). The Instigator affects the second prominent participant, the Patient, sending a cause-SIGNAL which triggers a transition of the Patient (cf. Figure 1). In the derived decausative reflexive, the Patient becomes the focus of the activity: it becomes the only prominent participant and its prototypical semantic role is transformed to that of an Experiencer who is affected by an action which can be generally defined as "happening by itself". Figure 6 provides the decausative EVENTITY FRAME.

Examples of decausatives are: *vbesja se* ('get enraged'), *vloša se* ('get worse'), *vdâhnovja se* ('feel inspired').

4. Discussion and further development

Table 1 represents the quantitative distribution of the analyzed data into the semantic classes outlined above.

Semantic Class	Number
Inherent	37
Motive	107
Absolute	35
Deaccusative	10
Decausative	120
Other	28

Table 1: Quantitative distribution of verbs

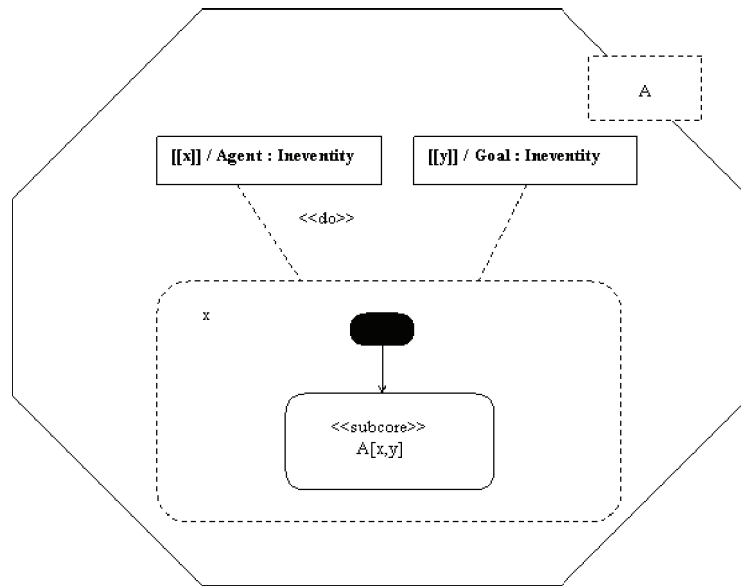


Figure 5: Deaccusative reflexives.

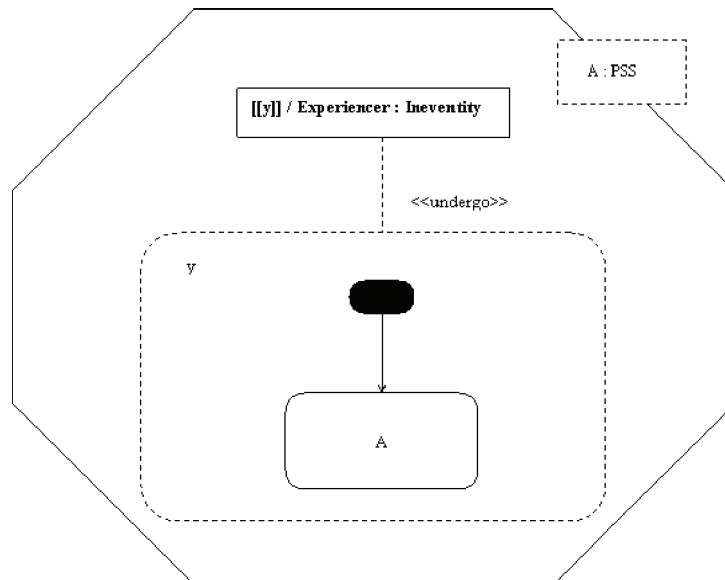


Figure 6: Decausative reflexives.

In some verb pairs the reflexive counterpart has a totally different meaning compared to the non-reflexive verb. Such reflexives (28 in number) fall out of the current classification.

The factors that influence the categorization of the reflexives are:

- lexical paradigmatic approach to the data;
- for each verb only one reading is chosen;
- the semantic descriptors are top level generalizations.

The UER formalism provides the possibility to capture variable granularity of the semantic description. The UER metamodel with its multiple layers of abstraction guarantees the deployment of a type hierarchy and makes use of

the distinction between type and instance. The generalization mechanism is a fundamental one and allows the user to adjust the granularity of the linguistic modeling depending on the application requirements. Thus the semantic descriptions of the classes of Bulgarian verbs, represented in this paper, are generalized in TEMPLATES, that is, they are parameterized EVENTIVITY FRAMES.

The next step is to build more specific descriptors by using an inventory of semantic primes compiled on the basis of semantic sources like Semantic Minimum Dictionary (Kasabov, 1990), and Semantic Language (Apresjan, 1974).

At the same time a cross-lingual Bulgarian - French investigation is carried out where the classified Bulgarian verbs are used as a seed data set.

The language models described in this paper provide the possibility for building linguistic components utilizable in knowledge-driven systems.

5. References

- Apresjan, J. D. (1974). *Lexical Semantics*. Nauka Publishing House, Moscow (in Russian).
- Boteva, S. (2000). *The Verb in French and Bulgarian. Functional-Semantic Grammar*. "Kolibri" Publishing House, Sofia.
- Genyushene, E. S., Nedyalkov, V. P. (1991). Typology of Reflexive Constructions. In *Functional Grammar Theory: Category of Person. Category of Voice*. Nauka Publishing House, Sankt-Petersburg, pp. 241-276 (in Russian).
- Kasabov, I. (1990). *Semantic Dictionary - Minimum*. Sofia University Publishing House "Kliment Ohridski", Sofia.
- Kordi, E. E. (1981). Derivational, Semantic and Syntactic Classification of French Pronominal Verbs. In *Verbal voice constructions in languages with diverse structure*, "Nauka" Publishing House, Leningrad, pp. 220-254 (in Russian).
- OMG Unified Modeling Language Specification. <http://www.omg.org/>
- Paskaleva, E., Simov, K., Damova, M., Slavcheva, M. (1993). The Long Journey from the Core to the Real Size of a Large LDB. In *Proceedings of ACL Workshop "Acquisition of Lexical Knowledge from Text"*, Columbus, Ohio, pp. 161-169.
- Paskaleva, E. (2003a). Automatic Processing of Bulgarian and Russian Language Resources in Consecutive and Parallel Mode. *Proceedings of the Conference Slavistics at the Beginning of 21 Century. Traditions and Expectations.*, SemaRSH Publishing House, Sofia, Bulgaria, pp. 111-120 (in Bulgarian).
- Paskaleva, E. (2003b). Compilation and Validation of Morphological Resources (Overview of the Morphology Cooking Technologies). In S. Piperidis and V. Karkaletsis, editors, *Proceedings of the International Workshop on Balkan Language Resources and Tools, First Balkan Conference on Informatics*, Thessaloniki, Greece, pp. 68-74.
- Schalley, A. C. (2004). *Cognitive Modeling and Verbal Semantics*. Mouton de Gruyter, Berlin - New York.
- Slavcheva, M. (2003a). Some Aspects of the Morphological Processing of Bulgarian. In *Proceedings of the Workshop on Morphological Processing of Slavic Languages, EACL 2003*, Budapest, Hungary, pp. 71-77.
- Slavcheva, M. (2003b). Integrating a Verb Lexicon into a Syntactic Treebank Production. In *Proceedings of the Second Workshop on Treebanks and Linguistic Theories*, Vaxjo, Sweden, pp. 165-176.
- Slavcheva, M. (2004). Verb Valency Descriptors for a Syntactic Treebank. In *Proceedings of LREC 2004*, Lisbon, Portugal, pp. 1153-1156.