Linking Reflexive Verb Structure to Verb Meaning in a Cross-Lingual Lexical Setting

Milena Slavcheva [1]
Bulgarian Academy of Sciences

Abstract

Language typology studies have shown that reflexive verb structures are widely represented in diverse languages. In this paper reflexive verb constructs in Bulgarian, French and Hungarian are compared and described on a paradigmatic level. A classification is provided where key Bulgarian reflexive verb constructs, distributed in semantic classes, are used as a seed data set for defining corresponding verb constructs in French and in Hungarian. Such a systematic semantic interpretation of reflexive constructs provides a considerable amount of linguistic knowledge about the possible relations between the predicates and their arguments. The semantic representational framework, used for building descriptors at a conceptual level, is the Unified Eventity Representation (UER), which provides adequate tools for overall structuring of semantic linguistic knowledge, unifying in an innovative way relevant approaches from linguistics and design principles from object-oriented programming.

Keywords : verb predicate, reflexive verb constructs, eventity frame, semantic classification, cross-lingual setting, lexical semantics.

1. Introduction

Language typology studies have shown that reflexive verb constructs are widely represented in diverse languages (Genyushene & Nedjalkov 1991; Dezső 1984). Taken in its entirety, there is striking similarity among languages in the diversity of senses encoded in reflexive forms. What needs to be explored for practical purposes is the actual distribution of senses, expressed by reflexive forms, in a given language. Such a systematic semantic interpretation of reflexive constructs provides a considerable amount of linguistic knowledge about the possible relations between the predicates and their arguments (Semecký & Podveský 2006; Cankov 1995; Slivkova 1995).

The semantic description of reflexives is linked to morphosyntactic classes of verbs in a large electronic lexical database of Bulgarian, which has been utilized in a number of applications in the field of Human Language Technologies (HLT) (Paskaleva *et al.* 1993; Slavcheva 2003a). Having used exhaustively the potentialities of the morphosyntactic knowledge for linguistic predictions on shallow processing layers (Slavcheva 2003b; Slavcheva 2004), there emerged the task of augmenting semantically the verb classes for content analysis purposes (Slavcheva 2006a). At the same time, the modeling of linguistic phenomena in a given language becomes more significant when carried out in a contrastive setup for the purpose of multilingual applications (Reinhart & Siloni 2005). Reflexive verb constructs in Bulgarian, French and Hungarian are cross-linguistically explored and described. Key Bulgarian reflexive verb constructs, distributed in semantic classes, are used as a seed data set for defining corresponding verb constructs in French and in Hungarian.

[1] Institute for Parallel Processing, Bulgarian Academy of Sciences, milena@lml.bas.bg

Currently, the contrastive modeling is carried out on a paradigmatic level. The source data are Bulgarian verb lexemes included in a frequency-based lexicon, which is a subset of the large lexical database. The frequency-based lexicon is compiled on the basis of a newspaper corpus. The semantically analysed constructs consist of a verb and a reflexive pronominal clitic in Bulgarian. The semantically categorized reflexive patterns in Bulgarian are contrasted with patterns in French and in Hungarian. The obtained Bulgarian-French and Bulgarian-Hungarian equivalents fall in two categories according to the correspondence between content and form: 1) one and the same meaning is expressed by what are considered corresponding forms in the language pairs; 2) one and the same meaning is expressed by different, not corresponding forms.

The semantic representational framework, used for building descriptors at a conceptual level, is the Unified Eventity Representation (UER), developed by Andrea Schalley (Schalley 2004). The UER, based on the Unified Modeling Language (UML) (OMG), is a graphical formalism, introducing the object-oriented system design to linguistic semantics. This framework provides adequate tools for overall structuring of semantic linguistic knowledge, unifying in an innovative way relevant approaches from linguistics and design principles reminiscent of a well-established up-to-date programming paradigm (cf. (Eckel 2002)).

The paper is structured as follows. In section 2 a classification of reflexives is introduced, taking into account the semantic properties of paradigmatic morphosyntactic patterns. Section 3 represents the linking of UER modeling elements to types of verb predicates. Section 4 provides the results of an initial experiment exploring the extent of formal equivalence of conceptually equivalent units in the language pairs. Section 5 briefly discusses related work and further development.

2. Classification-driven properties of reflexives

The issue of reflexive verb forms is significant due to the opposition with non-reflexive verb forms. In Bulgarian and French there are verb pairs where the verb forms coincide, but the difference is in the absence or presence of a clitic (e.g., bg. *izmâčvam* / *izmâčvam se*, fr. *torturer* / *se torturer*, *emmitoufler* / *s'emmitoufler*), while in Hungarian the verb pairs consist of words with the same root but opposing suffixes (or opposing null suffix and a reflexive suffix) positioned before the inflection (e.g., hu. *ideges-ít* (trans.) / *ideges-**ked**-ik* (refl.), *mos* (trans.) / *mos-**akod**-ik* (refl.)). An important indicator of reflexive structures are the full forms of the reflexive pronouns in the three languages in question (see the *Reflexive Marker 2* column of Table 1). The structures they form, however, are syntax-driven and that is why the current lexicon-driven investigation takes into account the full forms of reflexive pronouns only where necessary for the analysis. Table 1 summarizes the reflexive markers in Bulgarian, French and Hungarian.

Language	Reflexive Marker 1	Reflexive Marker 2
Bg.	se, si	sebe si
Fr.	me, te, se, nous, vous, se	(tonic pron)-même
Hu.	-ód(ik), -őd(ik), -kod(ik), -ked(ik), -köd(ik), -koz(ik), -kez(ik), -köz(ik), -ul, -ül, etc.	maga+inflection

Table 1. Reflexive markers

Table 2 represents a classification of reflexives according to the meaning shift triggered by the combination of a verb and a reflexive marker. It should be noted that the explored source construct is the combination of a verb and a reflexive pronominal clitic in Bulgarian. Thus Table 2 contains an exhaustive classification of the semantic types in Bulgarian linked to the morphosyntactic types in a paradigmatic lexically oriented setting. Each type label in Table 2 is provided with indications (in the line under the label) of the languages (i.e., bg, fr, hu) in which the given sense is expressed in a form containing the respective reflexive marker for each language (pronominal clitic for Bulgarian and French, suffix for Hungarian).

The *correlation* between reflexive and non-reflexive verb pairs can be viewed as the first filtering parameter that has two possible values indicating the existence (**correlative(+)**) or non-

correlative(-)	correlative(+)			
	CorrelativeIntrans		CorrelativeTrans	
	regularity(+)	regularity(-)	regularity(+)	regularity(-)
bg, fr, hu	inherent refl. (*bg, fr, hu*) reciprocals (*bg, fr, hu*) modal dative (*bg*) optative (*bg*) impersonal (*bg*)	formal (*bg, fr, hu*) different lexeme (*bg, fr, hu*) idiosyncratic (*bg, fr, hu*)	inherent refl. (*bg, fr, hu*) reciprocals (*bg, fr, hu*) passive (*bg, fr*) dative (*bg, fr*) modal dative (*bg*) optative (*bg*) impersonal (*bg*)	motive (*bg, fr, hu*) absolutive (*bg, fr, hu*) deaccusative (*bg, fr, hu*) decausative (*bg, fr, hu*)

Table 2. Classification of reflexives

existence (**correlative(-)**) of a non-reflexive correlative. Synchronically, the members of this verb group, usually known as *reflexiva tantum*, are considered arbitrary for each language (e.g., bg. *usmihvam se* (refl.) / fr. *sourire* (not refl.) / hu. *mosolyog* (not refl) 'smile'; fr. *s'absenter* (refl.) / bg. *otsâstvam* (not refl.) / hu. *hiányzik* (not refl.) 'be absent'; hu. *vonat-**koz**-ik* (refl.) / bg. *otnasjam se (do)* (refl.) / fr. *concerner* (not refl.) 'be related to'). Nevertheless there is coincidence of form and meaning among languages worthy to be explored. For instance, the current investigation has shown that around 33% of the compared Bulgarian correlative(-) verbs are also reflexive in French (e.g., bg. *provikvam se* / fr. *s'écrier* 'cry out'). For Hungarian the reflexive equivalents of the Bulgarian verbs are around 46% (e.g., bg. *gordeja se* / hu. *büszkél-ked-ik* 'be proud').

The reflexives that have a non-reflexive correlative are first classified according to the *transitivity* property of their non-reflexive correlative, which can be *intransitive* (**CorrelativeIntransitive**) or *transitive* (**CorrelativeTransitive**). The intransitive correlative can be a true intransitive verb or a verb usually having an indirect object complement. The transitivity property, stemming from morphosyntax, should be viewed in a conceptual perspective: the relation of arguments to their predicates within the event concepts, or *eventities* as defined in the UER cognitive approach (Schalley 2004).

In a paradigmatic perspective, *regularity* is an important parameter that determines the distribution of reflexives into two gross categories. Regularity is the possibility of "adding" a rather fixed sense. The result is sense accumulation, which is predictable and productive, that is, a number of verb predicates are "uniformly" affected by the sense addition. The **regularity(+)** data types are obtained by a general rule, while the **regularity(-)** data types are the result of special rules or are indicated individually.

The regularity parameter is related to the *degree of lexicalization* of the reflexive verb forms, which determines the arrangement of the modeling elements in the descriptors. If we imagine a scale of lexicalization, the regularity(+) types occupy its non-lexicalized, "grammatical" extreme slots, while the regularity(-) types approximate the lexicalized extremes of the scale.

There is also a scale of *remoteness* of the reflexive connotation expressed by the reflexive forms from what is considered the prototypical reflexive sense. For instance, the meaning of the regularity(-) reflexives is generally considered to be pseudo-reflexive in contrast to the meaning of inherent reflexives.

The inherent reflexives have the prototypical reflexive sense which is usually related to the idea of coreferential arguments: an action of an animate agent is directed towards himself. The idea of a causation relation is preserved, but it is realized between coinciding subject and object. Most of the representatives of inherent reflexives belong to the CorrelativeTransitive type, for example, bg. *mija se* / fr. *se laver* / hu. *mos-akod-ik* 'wash onself'; bg. *izmâčvam se* / fr. *se torturer* / hu. *gyötr-őd-ik* 'torment oneself'. The CorrelativeIntransitive type of inherent reflexives is represented by a small number of predicates which take a specific indirect object, for instance, bg. *protivorecha na njakogo*, literally traslated as 'contradict to someone' / *protivorecha si* 'contradict onself'.

In general, the CorrelativeIntransitive type of reflexives, bearing the feature *regularity(+)*, is rather language specific. The same is valid for the CorrelativeIntransitive reflexives defined as *regularity(-)*. It should be noted that as far as the formal reflexives are concerned, there is (almost) no change in meaning (e.g. bg. *beleja/beleja se* 'be white'), but they are defined as regularity(-) due to the fact that they are quite arbitrary and as such have to be individually indicated.

Most of the reflexives, defined as CorrelativeIntransitive, have different lexical meaning compared to the intransitive non-reflexive counterpart, hence they are identified as different lexemes. For example, bg. *ljutja* 'be peppery' / *ljutja se* 'go into a temper, flare up', *otivam* ('go, set out') / *otivam si* ('go home'); fr. *douter* ('doubt, question') / *se douter* ('suspect').

The reflexives with a transitive non-reflexive correlative (CorrelativeTransitive) are the most numerous and the most significant ones for the language phenomenon in question. Key semantic distinctions, relevant to them, are considered in section 3 in relation to the appropriate assembling of modeling elements. Here brief definitions and examples will be given for some classes of reflexives.

The regularity(-) type of reflexives can be generally named *figurative*, since in the majority of eventities they encode, the action is metonymically (in a broad sense) or metaphorically attributed to the subject participant.

The *motive* reflexives encode an eventity where a single participant is generically defined as the actor who's body is intrinsically and fully involved in the action. The movement sense is more or less figurative and refers to the physical or spiritual "self" of the participant (e.g., *nastanjavam se*, fr. *s'installer*, hu. *elhelyezkedik* ('settle oneself').

The *absolutive* reflexives are predominantly related to mental activities, activities of the will, social activities, etc. Prototypically the meaning of the absolutive reflexives is analogous to the meaning of absolutively used transitive verbs like *eat*, *read*, etc. (*He eats a sandwich.* / *He eats.*). Examples of absolutive reflexives are: bg. *izrazjavam se*, fr. *s'exprimer*, hu. *kifejeződik*, *megnyilvánul*; bg. *proiznasjam se*, fr. *se prononcer*, hu. *nyilatkozik*, *ejtődik* 'express oneself'.

In the case of *deaccusative* reflexives there is perspective shift between the agent and his goal: the status of the goal changes from that of a prominent participant in the eventity describing the initial, transitive, non-reflexive predicate to that of a non-prominent, but conceptualized participant in the eventity describing the derived reflexive predicate. Examples of deaccusatives are: bg. *približavam se*, fr. *s'approcher*, hu. *közeledik* 'get nearer to'.

The numerous *decausative* reflexives denote eventities where 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". Examples of decausatives are: bg. *vdâhnovjavam se*, fr. *s'inspirer, s enthousiasmer*, hu. *fellelkesül, föllelkesedik* 'feel inspired'; bg. *zabluždavam se*, fr. *se tromper*, hu. *eltéved* (the *-ed* suffix also has a sense of reflexivity) 'be misled'.

3. Model sets relevant to reflexivity

The UER framework introduces tools suitable for representing the many-sided phenomenon of reflexive verb structures denoting different senses at the borderline between the lexical and the syntactic level. The underlying cognitive approach of UER allows for building the semantic structures corresponding to the morphosyntactic structures without requiring strict isomorphism

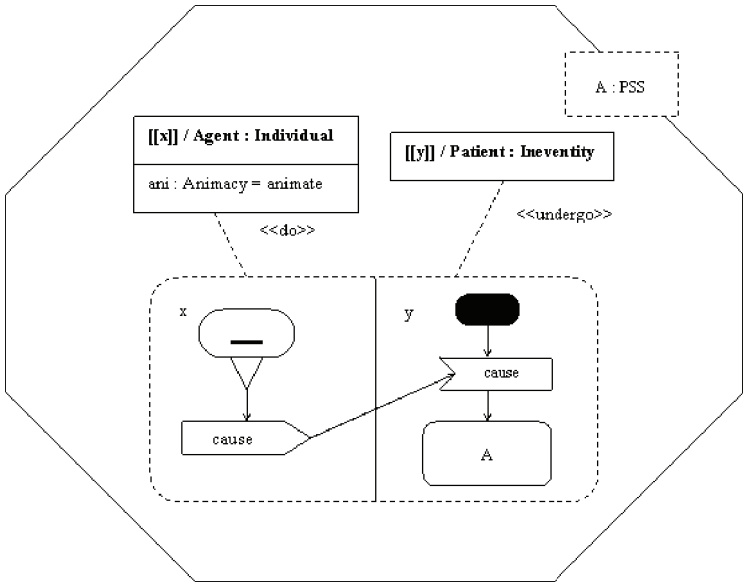

Figure 1. EVENTITY FRAME diagram of an initial transitive verb predicate.

between the semantic and the morphosyntactic categories. There is possibility for using intuition to a considerable extent, and the devices for structuring and management of the linguistic knowledge ensure the convenience of creating, adding and reusing linguistic components on demand. The relevance of the framework is also confirmed by the UER metamodel with its multiple layers of abstraction. The semantic description can be of variable granularity thanks to the fundamental generalization mechanism (Schalley 2004), allowing a user-defined degree of generalization or specification of the linguistic knowledge representation.

In order to build semantic descriptors for certain types of verb predicates, it is necessary to appropriately arrange the predefined modeling elements and to add the user-defined extensions, necessary for the current task. Due to space limits, only the most apparent differentiating properties of the types of verb predicates will be presented in this paper.

Figure 1 provides an idea of the UER modeling elements (Schalley 2004). An eventity is represented by an EVENTITY FRAME diagram (an octagon container), which includes a *dynamic core* and a *static periphery*. The dynamic core (the dashed outline rectangle with rounded corners) is a state chart depicting the state-transition system of the conceptualized actions. The static periphery (the rectangles in the upper part of the octagon) includes representation of the participants, their properties and relations. Figure 1 provides a TEMPLATE EVENTITY FRAME diagram (i.e., including parameters to be bound, indicated by the dash-outline rectangle in the upper right corner of the octagon) of a prototypical initial transitive verb predicate from which a majority of reflexives is derived. Its interpretation can be roughly formulated as "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: the agent, starting from an unspecified state triggers transition of the patient from an unknown state to some new state."

A key semantic feature is the presence/absence of *causation* relation, represented by the cause-SIGNAL modeling element (see Figure 1 for an illustration). The *inherent reflexives*, the *reciprocals* and the *datives* are considered to have the causation relation in their semantic descriptor. The subtyping of the regularity(+) reflexives is indicated by STEREOTYPES, like `<<reflexive>>`, `<<reciprocal>>`, etc., referring to the whole EVENTITY FRAME.

The rest of the semantic types in the classification are characterized by the absence of causation. For instance, Figure 2 provides a decausative EVENTITY FRAME TEMPLATE: there is one prominent participant, whose semantic role is Experiencer; the dynamic core depicts a transition, which is generally defined as "happening by itself" to the participant. Examples of decausatives are provided in section 2.

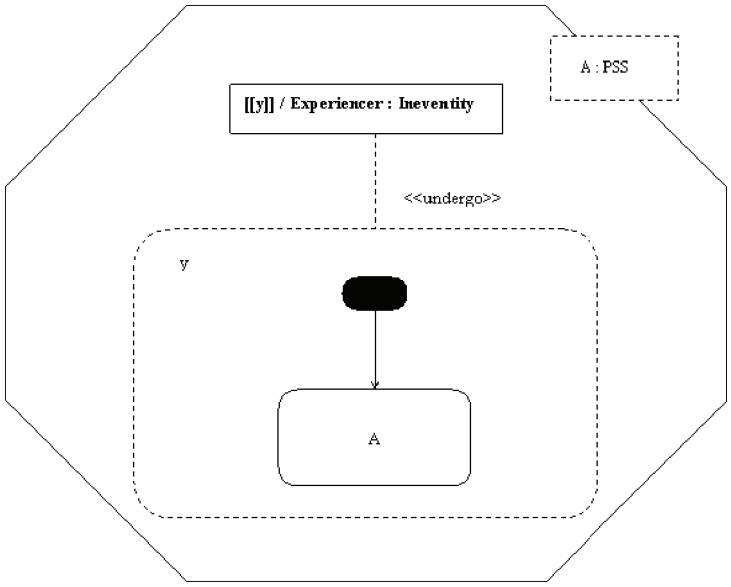

Figure 2. A decausative EVENTITY FRAME TEMPLATE

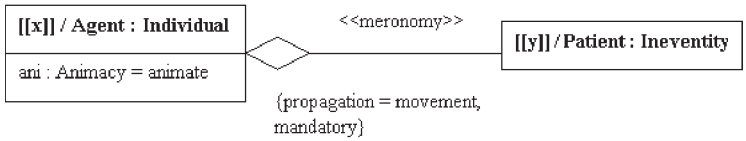

Figure 3. AGGREGATION relation of meronymy between PARTICIPANT CLASSES.

Another key semantic feature is the type of *association* relation between the participants. For instance, in the *partitive inherent* reflexives, the second participant (the object) is part of the first participant (the subject). In the UER formalism such a relation is named AGGREGATION and can be represented by the modeling element in Figure 3, which belongs to the static periphery of an EVENTITY FRAME diagram with two prominent participants (cf. Figure 1).

The AGGREGATION relation is specified as MERONYMY by the STEREOTYPE `<<meronymy>>` and is further specified by *secondary characteristics*, situated under the line, which connects the *determiner* (i.e., the left member of the relation) and the *tolerator* (i.e., the right member of the relation). Here are examples of a typical partitive reflexive construct in Bulgarian and French:

(1) Rita si mie râcete.
Rita refl wash-3p,sg hands-the.
'Rita is washing her hands.'

(2) Rita se lave les mains.
Rita refl wash-3p,sg the hands.
'Rita is washing her hands.'

The ATTRIBUTES are important modeling elements, which, as ENUMERATION classes, can specify the PARTICIPANT CLASSES, as well as the state-transition machine. For instance, they can represent semantic features like *animacy*, *humanness*, *volition*, *intentionality*, *fortuity*, *sentience*, relevant for describing the character of the eventity encoded by the verb predicate. For example, animacy is in many cases a differentiating feature between the inherent reflexive (the subject is an animate entity) and the passive sense (the subject is an inanimate entity). The representation of animacy as an ATTRIBUTE modeling element can be seen in Figure 1: 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.

4. Experiment

In the pilot data exploration carried out so far, 439 Bulgarian reflexive verb predicates have been provided with equivalents in French and Hungarian. In this way, Bulgarian-French and Bulgarian-Hungarian pairs have been obtained that contain conceptually equivalent units, which are also formally equivalent, that is, the eventities are encoded in a variant of a reflexive construct, specific for each language. Thus cross-lingual similarities and differences in the linking of conceptual structures to their formal expressions are registered. It should be noted that the represented in Table 3 reflexive verb units belong to semantic classes, which are close to the lexicalized extremes of the scale showing the degree of lexicalization. Table 3 provides the percentage of equivalence, calculated in relation to the Bulgarian reflexive verbs, taken as the basis.

Language	Inherent reflexives	Absolutives	Motives	Decausatives
Bg.	100%	100%	100%	100%
Fr.	80%	60%	70%	58%
Hu.	60%	36%	64%	79%

Table 3. Equivalence relation

The results are due to the typological deviation of Hungarian from Bulgarian and French. At the same time the high percentage of equivalent Hungarian decausatives is a very interesting result. It is related to the medium alternation, characteristic for this non-Indo-European language as well.

5. Related work and further development

The cross-lingual classification of reflexives represented in this paper is influenced by the typological classification of reflexive structures, proposed in (Genyushene & Nedjalkov 1991). The contrastive study of French and German reflexives in (Cortés & Kriegel 1997) has also been considered, as well as the analysis of the reflexivity expression in Hungarian in (Fóris-Ferenczi 2005).

The further development includes specification of the generic semantic descriptors using libraries of minimal sets of semantic primes (Slavcheva 2006b). The approach differs from that of the widely developed WordNets and VerbNets in the type and number of senses and sense relations: in WordNets and VerbNets the senses are very subtle and are of great number, organized by principles, mostly typical for lexicography. The UER eventity frames differ from the well-known FrameNets, where rather specific case frames are built in a sentence composition perspective. The descriptors built in the present work represent the minimally necessary arguments - those that are conceptualized in the eventity viewed from a decompositional semantic perspective (Apresjan 1974). At the same time WordNets, VerbNets, FrameNets could serve as valuable suppliers of semantic primes selected according to the demands of a given task.

An important investigation is related to the discovery of cross-lingual equivalents in parallel texts, where at the level of language encoding of concepts, paraphrasing would play the main role (Apresjan 1974), this time in a compositional (Pustejovsky 1995), text-driven semantic setting.

References

APRESJAN J. D. (1974), *Lexical Semantics*, Nauka Publishing House, Moscow, (in Russian).

CANKOV K. (1995), *Verbs and Verb Forms with a Reflexive Element in Modern Standard Bulgarian*, "St. Kliment Ohridski" University Publishing House, Sofia, (in Bulgarian).

CORTÉS C. et KRIEGEL S. (1997), "Étude traductologique (français-allemand) des constructions réfléchies dans des textes specialisés", in *Cahier du CIEL* : 157–194.

DEZSŐ L. (1984), *Typological Study of Russian Grammar in Contrast with Hungarian Grammar. Simple Sentence.*, Tankönyvkiadó, Budapest.

ECKEL B. (2002), *Thinking in Java, Third Edition*, Electornic Book, http://www.BruceEckel.com.

FÓRIS-FERENCZI R. (2005), *A visszahatás kifejesése a magyar nyelben történeti megközelítésben. (Reflexivity Expression in Hungarian in Historical Perspective.)*, Az Erdélyi Múzeum-Egyesület kiadása, Kolozsvár (Erdélyi Tudományos Füzetek, 251).

GENYUSHENE E. S. et NEDJALKOV V. P. (1991), "Typology of Reflexive Constructions", in *Functional Grammar Theory. Category of Person. Category of Voice*, Nauka Publishing House, Sankt-Petersburg : 241–276, (in Russian).

OMG, *Unified Modeling Language Specification*, http://www.omg.org/.

PASKALEVA E., SIMOV K., DAMOVA M. et 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* : 161–169.

PUSTEJOVSKY J. (1995), *The Generative Lexicon*, The MIT Press.

REINHART T. et SILONI T. (2005), "The Lexicon-Syntax Parameter: reflexivization and other arity operations", in *Linguistic Inquiry*, n^{o} 36, vol. 3.

SCHALLEY A. C. (2004), *Cognitive Modeling and Verbal Semantics*, Mouton de Gruyter, Berlin - New York (Trends in Linguistics).

SEMECKÝ J. et PODVESKÝ P. (2006), "Extensive Study on Automatic Verb Sense Disambiguation in Czech", in Sojka P., Kopeček I. & Pala K. (Eds), *TSD 2006*, Springer-Verlag, Berlin Heidelberg : 237–244 (Lecture Notes in Artificial Intelligence 4188).

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 : 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 : 165–176.

SLAVCHEVA M. (2004), "Verb Valency Descriptors for a Syntactic Treebank", in *Proceedings of LREC 2004*,Lisbon, Portugal : 1153–1156.

SLAVCHEVA M. (2006a), "Semantic Descriptors: The Case of Reflexive Verbs", in *Proceedings of LREC 2006*,Genoa, Italy.

SLAVCHEVA M. (2006b), "Semantic Representation of Events: Building a Semantic Primes Component", in Sojka P., Kopeček I. & Pala K. (Eds), *TSD 2006*, Springer-Verlag, Berlin Heidelberg : 245–252 (Lecture Notes in Artificial Intelligence 4188).

SLIVKOVA D. (1995), *The Reflexive Structures in Modern Bulgarian and German*, "St. Kliment Ohridski" University Publishing House, Sofia, (in Bulgarian).