Swedish-Czech Combinatorial Valency Lexicon of Predicate Nouns: Describing Event Structure in Support Verb Constructions

Silvie Cinková, Zdeněk Žabokrtský
Institute of Formal and Applied Linguistics,
Faculty of Mathematics and Physics, Charles University
Malostranské nám. 25, 118 00 Praha 1, Czech Republic
{cinkova,zabokrtsky}@ufal.mff.cuni.cz

Abstract

We have recently launched a project of an XML-based bilingual lexicon of predicate nouns. Besides itemizing the commonest support verb constructions and their Czech translation equivalents, the lexicon of predicate nouns aims at providing the user with relevant construction rules. It is mainly meant to help advanced Czech learners of Swedish to master delexicalized uses of the commonest lexical verbs in SVCs. It provides a systematic description of the valency of the nouns. Apart from that, it provides their support verb collocates, sorted by the Mel'čukian Lexical Functions.
A cross-linguistic application of the Transitivity Hypothesis is used in attempt to illustrate the lexical way of rendering event structure in Swedish, which poses problems for speakers of Czech, a language with morphological aspect. We believe that the morphosyntactic behavior of the noun together with telicity conditions affect the event structure of the entire SVC in context. The structure of the lexicon is theoretically based on the Functional Generative Description (FGD).

Introduction

This paper describes bilingual lexicographical processing of support verb constructions (SVCs) in a recently launched project of a machine-readable XML-based Swedish-Czech lexicon of predicate nouns. The lexicon is meant to capture delexicalized uses of the commonest lexical verbs in SVCs, in which the verbs show an evident tendency to grammaticalization (as defined by (Hopper, 1987) and further analyzed by (Heine, Claudi and Hünnemeyer, 1991)).

Support Verb Constructions, Support Verbs, Predicate Nouns

Support verb constructions are combinations of a lexical verb and a noun containing a predication. From the semantic point of view, the noun seems to be part of a complex predicate rather than the object (or subject) of the verb, despite what the surface syntax suggests. Support verbs are understood as verbs occurring in SVCs. Predicate nouns are in general nominal components of complex predicates (including SVCs).

An SVC is usually semantically transparent. Its meaning is concentrated in the noun phrase, while the semantic content of the verb is reduced or generalized. The matching verb is unpredictable, though often a metaphorical motivation can be traced back. Implicitly, SVCs affect the foreign language production rather than the reception (Heid, 1998), (Malmgren, 2002) and (Schroten, 2002).

If we look upon SVCs as collocations, the noun is the base, while the verb is the collocate; cf. e.g. (Malmgren, 2002), (Čermák, 2003) and (Schroten, 2002). Even in the cross-linguistic perspective it is the noun that constitutes the common denominator for equivalent support verb constructions, as empirically shown by (Fontenelle, 1992), whereas the support verbs do not necessarily match.

Important Features of the Swedish-Czech Combinatorial Valency Lexicon of Predicate Nouns

Besides itemizing the commonest SVCs and giving their Czech translation equivalents, the lexicon aims at providing the users with relevant SVC-construction rules for varying communication needs with special regard to event structure. The lexical evidence is always corpus-based.

Lemmatizing nouns both enables the enumeration of all verbs semantically related to the given noun together at one place and a more systematic description of restrictions in morphological number, article use and adjectival or pronominal modifications in the nouns. Inspired by (Hopper and Thompson, 1980), (Lindvall, 1998) and (Bjerre, 1999), we believe that morphosyntactic behavior of the noun together with lexical features of the support verb and of the event described by the predicate noun determine the event structure of the entire SVC employed in context.

Describing Valency in Predicate Nouns: Functional Generative Description

The lexicon displays the valency of the lemmatized predicate nouns within the FGD framework – a dependency-based formal stratificational language description framework that goes back to the functional-structural Prague School. For more detail see (Panevová, 1980) and (Sgall, Hajičová and Panevová, 1986). The theory of FGD has been implemented in the Prague Dependency Treebank project (Sgall, Panevová, Hajičová, 2004), a syntactically parsed corpus of Czech.

FGD can capture valency in the underlying syntax (the so-called *tectogrammatical language layer*). It enables listing of complementations (syntactically dependent autosemantic lexemes) in a valency lexicon, regardless of their surface (morphosyntactic) forms, providing them with semantic labels (*functors*) instead. It also regards coreference, ellipsis and topic-focus articulation. Implicitly, a complementation present in the tectogrammatical layer can either be directly rendered by the surface shape of the sentence, or it is omitted but can be inferred from the context or by common knowledge. A valency lexicon describes the valency patterns of a given lexeme (verb, noun, adjective or adverb) in form of

valency frames. In a valency lexicon the frames roughly correspond to lexical units in ordinary lexicons.

The lexicon of predicate nouns was significantly inspired by the closely related valency lexicons of Czech verbs VALLEX (though machine-readable, also designed for human use) and PDT-VALLEX (a supporting tool for treebank annotation, interlinked with the corpus data) – cf. (Straňáková-Lopatková et. al., 2002) and (Hajič et al., 2003). Though the lexicon of predicate nouns is primarily meant for human use, the quite rigid structure of (PDT-)VALLEX, whose both variants have originated from the needs of NLP-applications, seems to be helpful in remaining consistent when describing complex linguistic phenomena. On top of that, the Swedish part is very likely to prove useful in a possible annotation of an FGD-based Swedish treebank, when interlinked with the data in the same way as PDT-VALLEX.

Ordering Support Verbs under the Predicate Noun Lemmas: Lexical Functions

The lexicon of predicate nouns is also called a "combinatorial" one to show the acknowledgement for existing collocational dictionaries that have paid systematic attention to support verb constructions, e.g. (Benson, Benson and Ilson, 1986) and especially (Mel'čuk et al., 1984, 1992) and (Mel'čuk and Žholkovsky, 1984), modeling "institutionalized" lexical relations by the so-called Lexical Functions.

Lexical Functions are part of the Meaning-Text-Theory developed by Igor Mel'čuk and his collaborators (Mel'čuk, 1988), (Kahane, 2003). There are two elementary types of LFs – paradigmatic and syntagmatic – and this paper concerns only the latter. In terms of collocations, when two lexical units are collocates, one is usually the base that "selects" the other lexical unit to render a certain meaning together. The MTT captures it by the mathematical functional notation: LF_i (X) = Y, where X is called the keyword (the collocational base) and Y the value of the LF_i (the collocate). LFs can assign one value or a set of values to a given keyword. The values stand in the same lexical relation towards the keyword but they are not necessarily synonymous. The LFs describe the semantic relation between the keyword and the values. For examples and more details see (Wanner, 1996).

The following LFs are specific to SVCs; their keywords are the predicate nouns and their values are by definition verbs: **$Oper_1$**, **$Oper_2$, $Labor_{1,2}$, Copul** and **Func**.

In **Oper**, the predicate noun is a direct object of a transitive support verb, e.g. *pay attention*) or a prepositional object of an intransitive support verb, e.g. *get in touch*.

In **Labor**, the predicate noun is a prepositional object of a transitive verb, e.g. *subject sb to an interrogation*.

In **Copul**, the noun (or the adjective) is part of the predicate, in which a lexical verb has acquired a copula-like meaning, e.g. *fall ill. (= start to be ill)*.

In **Func**, the predicate noun is the subject of the verb, e.g. *The accusation came from John*.

(The example sentences originate from (Wanner, 1996) and (Macleod, 2002).) The numbers denote indexes of the complementations (participants) of the events described. No. 1 is the Actor, No. 2 is the Patient. When an LF is specified by 1, it means that the Actor of the verbal event is identical with the Actor of the event described by the noun. When an LF is specified by 2, it means that the Actor of the verbal event is identical with the Patient of the event described by the noun.

Entry Structure

On the topmost level, the lexicon is divided into word entries. Each word entry relates to one predicate noun lemma and its possible spelling variants. Homonyms get each an indexed word entry. Each entry comprises valency frames of the given predicate noun. The frames regard the noun simply as an abstract noun standing outside any SVCs. Fig. 1 shows two valency frames of the lemma *kritik*. The noun governs two complementations with functors. Their surface forms are also listed.

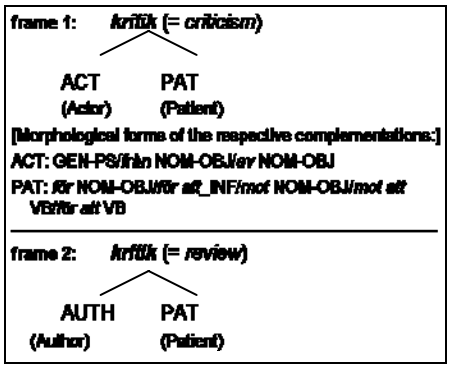

Fig.1: A word entry for kritik *(criticism) with two valency frames. The first frame includes the surface forms of the complementations by means of the SUC tagset (Ejerhed et al., 1992). Only the first frame renders the lexeme* kritik *as a predicate noun, the second frame shows* kritik *as an artefact, which cannot be a predicate noun.*

Each valency frame that renders a predicate noun (i.e. not the case of frame 2 in *kritik*) lists relevant SVCs, grouped according to LFs. These groups are technically called SVC-frames. An example is given by Fig. 2.

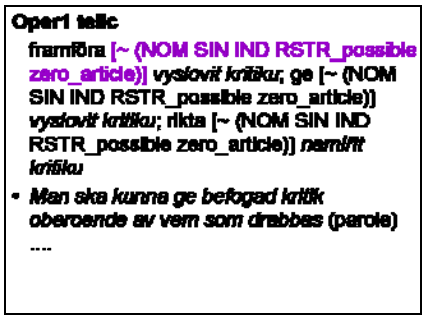

Fig. 2: One support verb construction frame nested in the first valency frame of kritik. *It is defined by the Lexical Function* **Oper$_1$**. *It includes telicity marking, description of morphosyntactic characteristics of the predicate noun* kritik *in combination with the support verbs* framföra, ge *and* rikta *(in square brackets after each verb), and Czech translation equivalents (in italics). Also an example sentence with reference to the PAROLE-corpus (http://spraakbanken.gu.se) is attached.*

Describing Event Structure in Swedish SVCs

SVCs are often referred to as one means of specifying event structure in non-aspectual languages as Swedish. A kind of event structure opposition is assumed between a SVC and its corresponding synthetic predicate (when there is any). Support verbs add further semantic features to the event described by the given predicate noun, such as inchoativity, durativity, terminativity and causativity (called *aspectual, diathetic* and *modal values* by (Fontenelle, 1992), or simply *aktionsart* by others, e.g. (Šmilauer, 1972)). However, this gives no direct correspondence to the Slavic category of aspect, which apparently is the product of more event structure features in combination, one of which being telicity. Also (Hopper and Thompson, 1980) emphasize the difference between aktionsart (which they call "lexical aspect", telicity and perfectivity (grammatical aspect).

Telicity, introducing the values "telic" and "atelic" should be regarded as independent of "aspect"/"perfectivity"/"boundedness" with its values "perfective" and "imperfective". More to this issue see (Nakhimovsky, 1996): *"A verb lexeme is telic if a simple declarative sentence in the past tense in which that lexeme is the main predicate is a telic sentence. A sentence is telic if it describes a telic process. A process is telic if it has a built-in terminal point that is reached in the normal course of events and beyond which the process cannot continue."* Nakhimovsky's claim that telicity is a lexical feature (i.e. semantically inherent to the verb in question) while aspect is inferred from semantico-syntactic relations in each given sentence, corresponds to (Pustejovsky, 1991), who, speaking of event-types, claims that *"the lexical specification of a verb's event-type can be overridden as a result of syntactic and semantic compositionality of the verb with other elements in the sentence"* and (Hopper and Thompson, 1980): *"Whereas telicity can be determined generally by a simple inspection of the predicate, perfectivity is a property that emerges only in discourse"*.

To summarize it, aktionsart and telicity are two different quantities, though they both are lexical features. Besides that, they both are to be discriminated from the grammatical aspect, whose morphological form they probably co-determine in aspectual languages as Czech.

The lexicon of predicate nouns captures aktionsart by attaching complementary LFs to the basic LFs listed above. It is phasal LFs – **Inc** (inceptive, inchoative), **Cont** (continuative) and **Fin** (finishing, terminative), causative LFs – **Caus** (causation), **Liqu** (causing to stop) and **Perm** (permitting to continue). Two more complementary LFs are employed, i.e. **Prox** (to be on the verge of) and **Anti** (negation). For details see (Wanner, 1996). The Anti-LF is mainly stated when the negation of the predicate noun is not allowed to negate the SVC and other means have to be used instead, such as the negation of the verb or using a support verb with the opposite meaning. The Anti-LF is not being stated consequently due to the lacking lexical evidence.

Issues of Telicity Marking in SVCs

It is to be stressed that SVCs are built as compositional events consisting of a "verbal" and a "nominal" subevent. Yet the "verbal" event does actually never "take place" due to the semantic depletion in support verbs (cf. (Fillmore, Johnson and Petruck, 2003)). The given support verb only passes some semantic features on to the "nominal" event. Durative events are by definition atelic (e.g. *have problems*), with the reservation that multiple telic "nominal" events combined with a durative atelic support verb express iterativity, e.g. *suffer from attacks*. (Below the "verbal" event corresponds to *subevent1* and the "nominal" event to *subevent*$_2$.)

SVCs denoting transitions (i.e. changes of state) are regarded as telic (cf. (Pustejovsky, 1991)), no matter what telicity value the given support verb would have if used as a lexical verb outside the SVC. This approach is based on (Bjerre, 1999). Bjerre puts it this way: "*SVCs denoting transitions are invariably achievements*[1], *either inchoatives or causatives [...], the SV always denotes an underspecified subevent*$_1$. *[...] Not surprising* terminative *is the negative counterpart of* inchoative. *[*Situationen kom ud af kontrol – (Situation_the came out of control)*] denotes a situation in which the resultant state is the negative of that in [*Situationen kom under kontrol = Situation_the came under control*] above. [...] This may be paraphrased: (subevent*$_1$*:) The situation was under control when something happened as a result of which (subevent*$_2$*:) the situation was out of (=not under) control"*. Bjerre notes that support verbs denoting transitions are either achievement verbs with inherently underspecified subevent$_1$ (*come, bring* etc.), or they are verbs of motion or location which lose their specific relation when used as support verbs.

[1] Transitions are further divided into two subtypes. In *achievements* the subevent$_1$ is underspecified, unlike in *accomplishments*, e.g. *Carl built a house* (accomplishment) × *The expedition reached the top of a mountain* (achievement). See (Bjerre, 1999).

For the purpose of the lexicon of predicate nouns, an SVC is thus marked as telic when:

a) both the subevent described by the predicate noun and the subevent described by the support verb are telic, e.g. *fatta beslut (take a decision)*

b) the subevent described by the support verb is atelic and the subevent described by the predicate noun is telic, e.g. *dra en slutsats (draw a conclusion)*

c) the subevent described by the support verb is telic and the subevent described by the predicate noun is atelic, e.g. *få besvär (get problems)*.

The event a) describes the termination of a process, and so does the event b) while the event c) describes the onset of a state, thus is inchoative (inceptive).

Perfectivity as a Transitivity Component

Our attempt to make a link between the Swedish and the Czech ways of specifying event structure is based on (Lindvall, 1998) and on (Lindvall, 2001), a summarizing article. Lindvall has performed a comprehensive parallel-corpora based comparison of Greek, Polish and Swedish to look into verbal boundedness and object definiteness as two interacting components of Transitivity. We make use of her inferences regarding Swedish and we assume that her inferences regarding Polish will also apply to Czech, as Czech and Polish are tightly related languages.

Lindvall's point of departure is (Hopper and Thompson, 1980). By comparison of many unrelated languages they analyze Transitivity, a universal linguistic phenomenon, intuitively understood as transfer of an activity from an Agent to a Patient, producing some effect. Hopper and Thompson isolate component parts/ parameters of the Transitivity notion with regard to the information structure of the given utterance, concluding that Transitivity is a continuum. Their parameters of Transitivity suggest each a scale according to which clauses can be ranked – see Fig. 3.

TRANSITIVITY:	**HIGH**	**LOW**
A. Participants	2 or more participants, A and O	1 participant
B. Kinesis	action	non-action
C. Aspect	telic	atelic
D. Punctuality	punctual	non-punctual
E. Volitionality	volitional	non-volitional
F. Affirmation	affirmative	negative
G. Mode	realis	irrealis
H. Agency	A high in potency	A low in potency
I. Affectedness of O	O totally affected	O not affected
J. Individuation of O	O highly individuated	O non-individuated

Fig. 3. Components of Transitivity proposed by Hopper and Thompson. The letter A *means Agent,* O *means Object.*

Hopper and Thompson further claim that the component features of Transitivity *"CO-VARY extensively and systematically [...] whenever an obligatory*

pairing of two Transitivity features occurs in the morphosyntax or semantics of a clause, THE PAIRED FEATURES ARE ALWAYS ON THE SAME SIDE OF THE HIGH-LOW TRANSITIVITY SCALE". They introduce the Transitivity Hypothesis: *"If two clauses (a) and (b) in a language differ in that (a) is higher in Transitivity according to any of the features A-J, then, if a concomitant grammatical or semantic difference appears elsewhere in the clause, that difference will also show (a) to be higher in Transitivity."*

Lindvall has proved that the Transitivity Hypothesis applies even cross-linguistically, having shown on Greek (a language employing both morphological aspect and noun definiteness) that utterances with high Transitivity tend to have perfective verb forms and definite objects, while utterances with low Transitivity tend to have imperfective verb forms and indefinite objects. Then she compared translations between Swedish (a noun-definiteness language) and Polish (an aspectual language) in both directions. It proved evident that in utterances with high Transitivity, Polish translations from Swedish tend to have perfective verb forms and Swedish translations from Polish tend to have definite noun forms, while low Transitivity utterances tend to have imperfective verb forms (Polish) and indefinite noun forms (Swedish). The observed noun definiteness was not confined to morphosyntactic features but resulted from the semantics of the noun phrase, which, on the other hand, was very often reflected by morphosyntax. This is why the lexicon of predicate nouns includes a detailed description of the morphosyntactic behavior of the predicate nouns in SVCs. For more details on the data structure of the lexicon see (Cinková and Žabokrtský, 2005).

A special feature of SVCs is that telicity is not determined by the verb but by the "nominal" event (yet modified by the support verb, cf. above). It is again the definiteness of the predicate noun that co-determines perfectivity. This will become apparent in selections of verb aspect forms in Czech translations of Swedish utterances and in noun definiteness in Swedish translations of Czech utterances. We assume that SVCs, especially those denoting transitions, have potentially rather high Transitivity also according to other parameters. Just to name a few, predicate nouns in SVCs characterized by all LFs except **Func** and **Copul** are morphosyntactic objects totally affected by the support verbs – the "nominal" events "come into existence" only by being named together with the given support verb – cf. the discussion of "effected objects" in (Barón and Herslund, 1998). Besides that, SVCs used as a means of transforming a state or a process into a transition imply discourse foregrounding. Yet the degree of Transitivity of an utterance in discourse shifts with other parameter values, especially with volitionality, affirmation and mode (see Fig. 3), which could explain the rather high morphosyntactic variation in predicate nouns captured in the lexicon.

Conclusion

The Swedish-Czech Combinatorial Valency Lexicon of Predicate nouns is an attempt to make use of the Transitivity Hypothesis, cross-linguistically applied by

(Lindvall, 1998), in order to describe the potential of event structure modifications in Swedish SVCs for Czech learners. The ultimate objective is to help Czech learners of Swedish with bridging the mental gap between an aspectual and a non-aspectual language by better understanding and active usage of the lexical mechanisms that affect event structure in Swedish SVCs.

Acknowledgements

This work has been kindly supported by GAUK 489-2004 and GAČR 405/04/0243.

References

I. Barón and M. Herslund. 1998. *Support Verb Constructions as Predicate Formation.* In " The Structure of the Lexicon in Functional Grammar". John Benjamins.

M. Benson and E. Benson and R. Ilson. 1986. *The BBI Combinatory Dictionary of English*.

T. Bjerre. 1999. *Event Structure and Support Verb Constructions.* In "Proceedings of the ESSLLI Student Session 1999.

S. Cinková and Z. Žabokrtský. 2005. *Treating Support Verb Constructions in a Lexicon: Swedish-Czech Combinatorial Valency Lexicon of Predicate Nouns.* In "Interdisciplinary Workshop on the Identification and Representation of Verb Features and Verb Classes – Saarland University". pages 22-27. Saarbrücken.

F. Čermák. 2003. *Abstract Nouns Collocations: Their Nature in a Parallel English-Czech Corpus.* In "Meaningful Texts: The Extraction of Semantic Information from Monolingual and Multilingual Corpora". Birmingham.

E. Ejerhed et al. 1992. *The Linguistic Annotation System of the Stockholm-Umeå Corpus Project, Version 4.31*. Publications from the Department of General Linguistics, University of Umeå, no. 32.

Ch. J. Fillmore, Ch. R. Johnson and M. R. L. Petruck. 2003. Background to FrameNet. *FrameNet and Frame Semantics*. *International Journal of Lexicography* (Special Issue, Guest Editor: T. Fontenelle)16: 235-250.

T. Fontenelle. 1992. *Cooccurrence Knowledge, Support Verbs and Machine Readable Dictionaries*. In "Proceedings of the 2nd International Conference on Computational Lexicography, COMPLEX'92, Budapest, Hungary". Linguistic Institute, Hungarian Academy of Sciences. pages 137-145. Budapest.

J. Hajič et al. 2003. PDT-VALLEX: *Creating a Large-coverage Valency Lexicon for Treebank Annotation.* In "Proceedings of The Second Workshop on Treebanks and Linguistic Theories. Växjö, Sweden, November 14 - 15, 2003". pages 57-68.Växjö.

U. Heid. 1998. *Towards a corpus-based dictionary of German noun-verb Collocations*. In "Actes EURALEX'98 Proceedings". pages 301-312. Liège.

B. Heine, U. Claudi and F. Hünnemeyer. 2001. *Grammaticalization. A conceptual framework*. Chicago.

P. Hopper. 1987. Emergent Grammar. *BLS,* 13:139-157.
P. Hopper and S. A. Thompson. 1980. Transitivity in Grammar and Discourse. *Language*, 56:251-299.
S. Kahane. 2003. *The Meaning-Text Theory.* In "Dependency and Valency. An International Handbook on Contemporary Research". Berlin.
A. Lindvall. 1998. *Transitivity in Discourse. A Comparison of Greek, Polish and Swedish.* Lund.
A. Lindvall. 2001. Grund, aspekt och definithet – en studie i morfologi i grekiska, polska och svenska. In: "Postskriptum. Språkliga studier till minnet av Elsie Wijk-Andersson". pages 170-184. Uppsala.
C. Macleod. 2002. *Lexical Annotation for Multi-word Entries Containing Nominalizations*. In "Proceedings of Third International Conference on Language Resources and Evaluation (LREC 2002); Las Palmas, Canary Islands, Spain". pages 943-948.
S.-G. Malmgren. 2002. Begå *eller* ta självmord*? Om svenska kollokationer och deras förändringsbenägenhet 1800-2000.* Göteborg.
I. A. Mel'čuk et al. 1984, 1992. *Dictionnaire explicatif et combinatoire du français contemporain*, Volume I and II. Montreal.
I. A. Mel'čuk. 1988. *Dependency Syntax: Theory and Practice.* New York.
I. A. Mel'čuk and A.K. Žholkovsky. 1984. *Tolkovo-kombinatornyj slovar' sovremennogo russkogo jazyka*. In "Wiener Slawistischer Almanach. Sonderband 14." Vienna.
A. Nakhimovsky. 1996. *A Case of Aspectual Polysemy, with Implications for Lexical Functions.* In: "Lexical Functions in Lexicography and Natural Language Processing". Studies in Language Companion Series (SLCS), Vol. 31. pages 169-179. Amsterdam-Philadelphia.
J. Panevová. 1980. *Formy a funkce ve stavbě české věty.* Praha.
J. Pustejovsky. 2000. *Syntagmatic Processes*. In "Handbook of Lexicology and Lexicography". de Gruyter.
J. Pustejovsky. 1991. The Syntax of Event Structure. *Cognition*, 41:47-81.
J. Schroten. 2002. *Light Verb Constructions in bilingual dictionaries.* In "From Lexicology to Lexicography". pages 83-94. Utrecht.
P. Sgall, J. Panevová and E. Hajičová. 2004. *Deep Syntactic Annotation: Tectogrammatical Representation and Beyond.* In "Frontiers in Corpus Annotation. Proceedings of the Workshop of the HLT/NAACL Conference", pages 32-38.
P. Sgall, E. Hajičová and J. Panevová. 1986. *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects.* Dordrecht, Prague.
M. Straňáková-Lopatková et. al. 2002. *Valency Dictionary of Czech Verbs: Complex Tectogrammatical Annotation.* In "LREC2002, Proceedings, vol. III. ELRA, pages 949-956.
V. Šmilauer. 1972. *Nauka o českém jazyku.* Praha.
L. Wanner. 1996. (Ed.) *Lexical Functions in Lexicography and Natural Language Processing*. Studies in Language Companion Series (SLCS), Vol. 31. Amsterdam-Philadelphia.