

- a) postup prací při řešení grantového projektu/ rozbor řešení grantového projektu na pracovišti řešitele a spoluřešitelů včetně zdůvodnění případného odklonu řešení projektu proti schválenému návrhu projektu, dosažené výsledky projektu (u mezinárodních projektů včetně mezinárodní spolupráce):

The main goal of the project was to propose an adequate framework for the theoretical description of language phenomena at the syntax-semantics interface; the proposed framework has been based on the division between a lexical and a grammar component. This framework has been applied to lexicographic processing of language data. A close interplay between the theoretical research and its application to extensive data annotation has proved to represent a fruitful strategy for enhancing both sides involved.

During the project, all research areas identified in the proposal were surveyed (in accordance with the original project description and the project timetable as it has been designed and further specified in the project year reports).

#### 1 Lexicographic representation of various changes in valency structure of verbs

##### 1.1 Theoretical research; design of a formal model of lexicographic description

MFF UK: The main focus was laid on the refinement of the formal model of lexicographic representation of changes in the valency structure of Czech verbs [16-3], [16-2]. We proposed a division of the lexicon into two parts and evaluated theoretical adequacy and economy of such division, as well as the suitability of different types of rules for different types of Czech alternations [12-4]. We distinguished two types of alternations (changes in valency frames of verbs) on the basis of used linguistic means: (i) grammaticalized alternations (diatheses representing the central type) [16-2] and (ii) lexicalized alternations (also referred to as semantic diatheses) [13-1], [14-4]. Later on, our attention shifted from central changes to peripheral ones which had not been analyzed so far, esp. to changes associated with syntactic reflexivity [14-1], [14-9], [15-1], deletions [15-4], [15-5], and reciprocity.

The theoretical results were summarized in an extended study forming the theoretical part of the book Valenční slovník českých sloves VALLEX [Valency lexicon of Czech Verbs VALLEX] [16-1], namely Part I, Valence sloves a její zachycení ve slovníku VALLEX [Valency of Verbs and Its Description in the VALLEX Lexicon], as well as in Part III, Gramatická komponenta [Grammar component].

These theoretical aspects were projected onto the data component and incorporated in the VALLEX lexicon (as described in the following sections, <http://ufal.mff.cuni.cz/vallex/3.0/>).

SI CAS: The proposed model was also examined with respect to the possibility of describing changes in valency structures in Russian and Polish – the formal model of the lexicon has been enriched so that it covers both monolingual and multilingual data as well as their interlinking [14-8], [15-2]. Main problems were identified and their solutions were proposed, we dealt esp. with problems concerning different types of reflexive constructions [12-7], [16-4], problems with variants of reflexive/non-reflexive verbs [16-4], and with the reciprocity diathesis [15-11]; further, stylistic equivalents, acceptability of Internet examples, formal realization of participants (sets of prepositions, conjunctions, participles etc.), as well as specific properties of verbs of communication were studied [14-13].

##### 1.2 Grammatical and syntactic diatheses: theoretical and practical aspects

MFF UK: Grammatical changes in the valency structure of Czech verbs – diatheses, reflexivity and reciprocity – were studied and their representation by means of formal syntactic rules was elaborated [12-4]. The preliminary version of the rules was tested on the example of passive diathesis and reciprocity involving two actants [12-3]. These preliminary rules have been extended to cover all types of diatheses [16-1] as well as changes related to syntactic reflexivity and reciprocity [14-1], [15-1]. A special attention has been paid to the rules describing the surface form of the marked members of diatheses of verbs with a genitive complement [16-2]. The proposed syntactic rules were further tested, refined, and evaluated with respect to the corpus data. An inventory of these rules forms the grammatical component of the lexicon, which is available both as a machine readable file and as part of the printed edition of VALLEX [16-1], where it has been extended with explanatory examples.

As for practical outputs, a tool for deriving valency frames for marked uses of verbs from the unmarked use (i.e., the active use) was designed.

Further, a system for the automatic identification of lexical units of Czech verbs that enter a particular type of diathesis was elaborated and tested on the recipient passive diathesis [13-4], [14-7], [16-2]; the automatic preprocessing was followed by a manual annotation (see also sect. 1.6).

All the theoretical results were incorporated in the lexicon data [12-6], [15-8], [16-1] (see Section 1.6).

SI CAS: The formal model comprising multilingual data was used as a basis for the annotation of grammatical diatheses in Russian and Polish: the model was tested on a basic group of Russian and Polish verbs (that enter lexicalized alternations, see 1.3); these verbs were annotated with respect to their diatheses (passive, deagentive, resultative) and reciprocity. As a result, it appeared that a set of values of the attribute diat proposed for Czech verbs

is not sufficient for Polish in case of deagentive diathesis and it should be enriched with further values [12-7]. A detailed description of Czech resultative constructions was given in [15-3].

Further, a comparison of reciprocity in Russian and Polish on one side and in Czech on the other was carried out and presented in [15-11]. And finally, Polish reflexive constructions were compared with Czech ones [16-4].

### 1.3 Lexicalized alternations (semantic diatheses): theoretical and practical aspects

MFF UK: The formal model of the representation of lexicalized alternations (i.e., relations between different lexical units of the same lexeme that are characterized by the same situational content but have different syntactic structures) was proposed in the doctoral thesis [12-5]; the thesis served as a basis for the monograph on conversive lexicalized alternations [14-3]. Further, a detailed analysis of three basic types of lexicalized alternations was carried out [13-1], [14-4]: (i) While lexical semantic conversions represent the central type of lexicalized alternations in the language system, as e.g. locative diathesis (e.g., ‘to load hay onto the truck’ – ‘to load the truck with hay’) and the swarm diathesis (e.g., ‘bees are swarming in the garden’ – ‘the garden is swarming with bees’) [12-2], (ii) the different structural expressions of a single situational participant, and (iii) the structural splitting of a situational participant are of a non-conversive character and are rather peripheral. Both conversive and non-conversive lexicalized changes were extensively documented by corpus evidence and their inventory was completed.

The information on all types of lexicalized alternations was integrated into the data component of the lexicon.

Moreover, a formal framework that allows us to represent different types of lexicalized alternations in the valency lexicon was proposed: the lexical rules describing changes in the correspondence between situational participants and valency complementations of verbs were formulated and projected to the grammar part of the lexicon [15-8], [16-1]

Part III, Gramatická komponenta [Grammar component], where it has been extended with explanatory examples.

SI CAS: As for the comparative part of the project, our research focused on a comparative analysis of lexicalized alternations in Russian and Polish and their possible combinations with grammatical diatheses. The original classification of lexical-semantic conversions (type (i)) formulated for Czech was proved to be relevant also for Polish and Russian verbs. The relevant types of lexical-semantic conversions were assigned to Russian and Polish counterparts of Czech lexical units entering these conversions [14-8], [14-10]. Our study revealed that lexicalized alternations – as they were identified for Czech – can be identified also in Russian and Polish (with the only exception of the so called “Theme – Path” alternation, which was not confirmed for Russian and Polish).

Further, Russian and Polish verbs with non-conversive lexicalized alternations, i.e. verbs allowing for the multiple structural expression of a participant and the structural splitting, were put under scrutiny. Several types of the multiple structural expression of a situational participant (type (ii), namely those concerning the Goal, Source, and Subject Matter) have been attested only with Czech verbs but not with their Russian and Polish counterparts [14-2]; other types (e.g. those involving the situational participant Area) were found only in Czech and Polish, but not in Russian. (iii) The results of the study focused on the structural splitting of a situational participant with Russian, Czech and Polish verbs of perception and verbs of communication is summarized in [15-2] and [14-13].

### 1.4 Other types of changes in valency structure of verbs

(In the project proposal, Section 1.4 dealt with non-conversive lexicalized alternations, the results are covered in Section 1.3 (types (ii) and (iii) above.)

MFF UK: A preliminary investigation revealed yet another phenomenon closely related to changes in the valency structure of Czech verbs, namely light verb constructions (in Czech linguistics usually referred to as analytical predicates). We made an initial step to a new formal lexicographic representation of these structures based on the decomposition of the valency information between (a) verbal valency frames and (b) nominal valency frames.

According to our proposal, both deep and surface syntactic structures of light verb constructions can be derived from the information given in the verbal and nominal frames, and three other special attributes by application of formal rules [13-2], [13-3], [14-11], [14-12].

An interplay of the formal rules for diatheses and rules determining formation of light verb constructions was studied as well [15-6].

### 1.5 Comparative aspects of diatheses

The comparative aspects of grammaticalized as well as lexicalized alternations were studied mainly by the SI CAS team. As a result of this research, several case studies were prepared, as was already mentioned in preceding sections.

As a result of a comparative study of Czech, Russian and Polish verbs focused on grammaticalized alternations (Section 1.2 above), a set of values of the attribute diat (originally proposed for Czech verbs) was enriched to cover also constructions specific for Russian and Polish [12-7]. Further, these three languages were compared with respect to the possibility to create reciprocal structures [15-11]; reflexivity in Czech and Polish were compared as well [16-4]. Second, different cases of lexicalized alternations in Czech, Russian and Polish were examined and compared. We focused esp. on the relation between syntactic and lexical expression of valency complementations of verbs; we

concluded that the lexical expression of valency complementations is restricted to some extent, the degree of such restrictions depends on a type of these combinations (idioms, set phrases etc. were examined especially in Czech, sometimes in comparison with Polish and Russian). Further, different types of alternations of Russian and Polish verbs were examined (lexicalized conversions and non-conversive lexicalized alternations [14-8], [14-10], [14-13] and [15-2]) – they were studied from the theoretical as well as practical point of view, as was summarized above in Section 1.3.

In addition, a case study of several Slovenian verbs (in comparison with their Czech counterparts in VALLEX) was initiated [13-5].

Finally, an attempt to exploit the VALLEX lexicon for automatic derivation of valency characteristics of related languages (for the purposes of NLP) was reported [14-6].

#### 1.6 Application in an electronic lexical resource

MFF UK: The results of the research carried out in all subtasks were applied in the VALLEX lexicon.

The new version of the VALLEX lexicon enhanced (i) with the grammar component specifying especially grammatical aspects of individual grammaticalized as well as lexicalized alternations, and (ii) with the information on applicability of individual alternations for individual lexical units in the data component.

As for (i), grammatical changes in the valency structure of Czech verbs related to diatheses, reflexivity and reciprocity were described in the form of formal rules. Further, changes in the correspondence between situational participants and valency complementations of verbs were described by lexical rules. These two types of rules form the grammatical component of the lexicon, which is available both as a machine readable file with syntactic rules and as a part of the printed edition of VALLEX [16-1], where it has been extended with explanatory examples.

As for (ii), a semi-automatic method allowing us to decide whether a certain diathesis is applicable to individual lexical units of verbs stored in the data component of the lexicon was formulated (see sect. 1.2). Its outputs were refined and manually corrected (if necessary) on the basis of a large amounts of corpus data [13-4], [14-7], [16-2]. The annotation of lexicalized alternations was done manually.

The final version of the on-line Czech valency lexicon, VALLEX, version 3.0, was released [15-8]. Both data and grammar component were enhanced with the information on diatheses, reciprocity and lexical alternations. A large set of automatic checks was run in order to prevent any inconsistencies; further, extensively data consistency checking was manually conducted.

VALLEX is available through a brand-new web interface <http://ufal.mff.cuni.cz/vallex/3.0/>. A printed edition was also published in accordance with the project proposal [16-1].

SI CAS: The results of studies focusing on comparative aspects were projected to the parallel annotation of selected Czech, Polish and Russian verbs. The annotators used an editor developed at MFF UK, which had been tailored for parallel annotation. As a result, the detailed annotation of selected groups of Russian and Polish verbs was prepared. Evidence from the Russian National Corpus ([www.ruscorpora.ru](http://www.ruscorpora.ru)), from the Polish National Corpus ([www.nkjp.pl](http://www.nkjp.pl)) and from Internet texts was used and corrected by native speakers; however, as some of the studied constructions are very rare, we had to rely also on introspective examples produced by native speakers.

The data were compared with the new version of VALLEX 3.0 and made available for further research at the departments involved in the project.

## 2 Mapping lexical resources: an effective way of enriching lexical information

A minor but still important task of the project consisted in enhancing the Czech valency lexicon with additional semantic information and with corpus evidence. This part of the project was conducted at MFF UK.

### 2.1 Enhancing Czech valency lexicon with semantic classes and semantic roles

For the purpose of enhancing the VALLEX lexicon with semantic information, we make use of the FrameNet project (<http://framenet.icsi.berkeley.edu>). In the first stage, semantic frames from FrameNet have been manually mapped onto eight groups of verbs with various semantic and syntactic properties. The feasibility of this task was verified by the achieved inter-annotator agreement measured on two semantically and syntactically different groups of verbs – verbs of communication and verbs of exchange (85.9% and 78.5%, respectively). Based on the upper level semantic frames from the relation 'Inheritance' built in FrameNet, the verbs of these eight groups were classified into more coherent semantic classes. Moreover, frame elements from these upper level semantic frames were assigned to valency complementations of the verbs of the listed groups as semantic roles. As in case of semantic frames, the achieved inter-annotator agreement in assigning frame elements measured on verbs of communication and verbs of exchange was promising (reaching 95.6% and 91.2%, respectively).

As a result, 1 270 lexical units belonging to the verbs of communication, mental action, psych verbs, social interaction, verbs of exchange, motion, transport and location (2 129 Czech verbs in total if perfective and imperfective verbs

being counted separately) were classified into syntactically and semantically coherent classes and their valency complementations have been characterized by semantic roles adopted from the FrameNet lexical database [12-1]. In this task, we originally planned to experiment with other groups of verbs and to increase the coverage of semantic information following the progress made in FrameNet. As no new version of FrameNet was released, adding the extracted semantic information to the HTML version of the lexicon was postponed till more data are available. Based on the results of the manual experiment, we prepared a semi-automatic method of mapping the information from FrameNet onto VALLEX [15-7]. However, despite promising results, a detailed manual analysis showed that the gained accuracy of the output is not satisfactory enough; thus this information is not included in the final version of VALLEX 3.0.

## 2.2 Strengthening the lexical resource with corpus evidence

The MFF UK team prepared a rule-based system linking lexicon entries across different versions of VALLEX. In this system, we have adopted an approach preferring high precision (reaching 91%) over coverage as it ensures high reliability of the obtained linked data. Thus system enables us to retrieve information from the previous versions of the lexicon, especially a corpus evidence from the VALEVAL corpus. The enriched lexicon, VALLEX version 2.6, was released for research purposes [12-6], see also <http://ufal.mff.cuni.cz/vallex/2.6/>.

This system was further elaborated and served as a basis for interlinking lexical units of two lexicons, VALLEX and PDT-Vallex (the valency lexicon of verbs from the Prague Dependency Treebank (PDT) and the Czech part of the Prague Czech-English Dependency Treebank (PCEDT)) [14-5], [15-7], [15-9]. This system has allowed us to enrich more than 2.5 thousand lexical units in VALLEX with a total of 75 thousand examples extracted from the PDT and PCEDT corpora, these examples are available in the on-line version of the lexicon [15-8].

The automatic links to PDT-Vallex on the level of individual lexical units were also exploited to add examples of the marked members of the passive diathesis for some verbs. Additional corpus examples were selected during the manual stage of the annotation [16-2].

(Since the CzEngVallex – an external resource, not part of this project – was published as late as in October 2015, the information obtained from this lexicon could not be included in VALLEX as was planned in the project proposal.)

The amount of research conducted by the MFF UK team as well as by the SI CAS research team fully corresponds to the percentage proportion of the partners on the project.

## Bibliography (all years, published, sorted by category, author(s), and year)

### Jimp:

[15-1] Kettnerová Václava, Lopatková Markéta, Panevová Jarmila: Shoda doplňku v reflexivních konstrukcích v češtině. Slovo a slovesnost, Vol. 76, No. 3, pp. 198-214, 2015. [15-1-SaS-reflexivity.pdf]

### Jneimp:

[12-1] Kettnerová Václava, Lopatková Markéta, Bejček Eduard: Mapping Semantic Information from FrameNet onto VALLEX. The Prague Bulletin of Mathematical Linguistics, No. 97, pp. 23-41, 2012. [12-1-PBML-sep.pdf]

[14-1] Kettnerová Václava, Lopatková Markéta, Panevová Jarmila: An Interplay between Valency Information and Reflexivity. The Prague Bulletin of Mathematical Linguistics, No. 102, pp. 105-126, 2014. [14-1-PBML-refl-sep.pdf]

[15-2] Skwarska Karolína.: Opis leksykograficzny syntaktycznych i semantycznych cech czasowników w języku polskim i rosyjskim na podstawie korpusów narodowych. Prace Filologiczne. Vol. LXVII, pp. 349-366, 2015. [15-2-Prace\_Filologiczne.pdf]

### Jrec:

[12-2] Kettnerová Václava: Syntaktické konstrukce typu Včely se hemží na zahradě – Zahrada se hemží včelami. Korpus – gramatika – axiologie, No. 5, pp. 54-71, 2012. [12-2-KGA-sep.pdf]

[14-2] Podobová Marie: Dynamické a statické vyjádření lokace v češtině a v ruštině. Nová čeština doma a ve světě, 1-2/2013, pp. 94-109, 2013. [14-2-Cestina-DIR-LOC-sep.pdf]

[15-3] Podobová Marie: Bezobjektové posesivní resultativní konstrukce od sloves tranzitivních a bezobjektové i objektové posesivní resultativní konstrukce od sloves intransitivních – jejich tvorba a funkce v českém jazyce. Nová čeština doma a ve světě, 1/2015, pp. 82-113, 2015. [15-3-Cestina-pos-res.pdf]

### B-book:

[14-3] Kettnerová Václava: Lexikálně-sémantické konverze ve valenčním slovníku. Karolinum, Prague, 282 pp., 2014.

[16-1] Lopatková Markéta, Kettnerová Václava, Bejček Eduard, Vernerová Anna, Žabokrtský Zdeněk: Valenční slovník českých sloves VALLEX. Karolinum, Prague, 700 pp., 2016.

### C-chapter in book:

[14-4] Kettnerová, V., Lopatková, M.: Alternace slovesných rámců. Kapitola v knize Jarmila Panevová a kol. *Mluvnice současné češtiny 2 / Syntax češtiny na základě anotovaného korpusu*. Praha: Karolinum, pp. 117-133, 2014. [14-4-Mluvnice-alter-sep.pdf]

[16-2] Lopatková Markéta, Vernerová Anna, Kettnerová Václava: Diateze ve Valenčním slovníku českých sloves VALLEX. In Skwarska, K., Kaczmarska, E. (eds.) *Výzkum slovesné valence ve slovanských zemích*. Slovanský ústav AV ČR, v.v.i., Prague, pp. 149-167, 2016. [16-2-kapitola-diateze-sep.pdf]

[16-3] Panevová Jarmila: Valence v gramatice, valence ve slovníku. In Skwarska, K., Kaczmarska, E. (eds.) *Výzkum slovesné valence ve slovanských zemích*. Slovanský ústav AV ČR, v.v.i., Prague, pp. 13-25, 2016. [16-3-kapitola-panevova.pdf]

[16-4] Skwarska Karolína, Bolbot Katarzyna, Filas Adriana: Zpracování polských zvrtných konstrukcí ve slovníku VALLEX. In Skwarska, K., Kaczmarska, E. (eds.) *Výzkum slovesné valence ve slovanských zemích*. Slovanský ústav AV ČR, v.v.i., Prague, pp. 169-180, 2016. [16-4-kapitola-skwarska.pdf]

D – proceedings in WoS:

[14-5] Bejček Eduard, Kettnerová Václava, Lopatková Markéta: Automatic Mapping Lexical Resources: A Lexical Unit as the Keystone. In *LREC 2014: Proceedings of the 9th International Conference on Language Resources and Evaluation*, European Language Resources Association, Reykjavík, Iceland, pp. 2826-2832, 2014. [14-5-LREC-vallink-sep.pdf]

[15-4] Hajič Jan, Hajičová Eva, Mikulová Marie, Mírovský Jiří, Panevová Jarmila, Zeman Daniel: Deletions and node reconstructions in a dependency-based multilevel annotation scheme. In *CICLing 2015: 16th International Conference on Computational Linguistics and Intelligent Text Processing*, LNCS 9041 Springer, Berlin / Heidelberg, pp. 17-31, 2015. [15-4-CicLing-sep.pdf]

[12-3] Kettnerová Václava, Lopatková Markéta, Uřešová Zdeňka: The Rule-Based Approach to Czech Grammaticalized Alternations. In *TSD 2012: Proceedings of the 15th International Conference Text, Speech and Dialogue*. In LNCS 7499, Springer Verlag, Berlin / Heidelberg, pp. 158-165, 2012. [12-3-TSD-gramalter-sep.pdf]

[14-6] Klyueva Natalia, Kuboň Vladislav: Automatic Valency Derivation for Related Languages. In *FLAIRS 2014: Proceedings of the Twenty-Seventh International Florida Artificial Intelligence Research Society Conference*, AAAI Press, Palo Alto, California, pp. 437-442, 2014. [14-6-FLAIRS\_czru-sep.pdf]

[14-7] Vernerová Anna, Kettnerová Václava, Lopatková Markéta: To pay or to get paid: Enriching a Valency Lexicon with Diatheses. In *LREC 2014: Proceedings of the 9th International Conference on Language Resources and Evaluation*, European Language Resources Association, Reykjavík, Iceland, pp. 2452-2459, 2014. [14-7-LREC-diateze-sep.pdf]

other journal articles:

[13-1] Kettnerová Václava, Lopatková Markéta: Lexikalizované alternace v češtině. *Linguistica Copernicana*, Vol. 9, No. 1, pp. 31-64, 2013. [13-1-Ling-Copernicana.pdf]

[14-8] Skwarska Karolína: Sintaksicheskiye i semanticheskiye xarakteristiki russkix, pol'skix i cheshskix glagolov v slovare sochetajemostej VALLEX. (Syntactic and semantic properties of Russian, Polish and Czech verbs in valency lexicon VALLEX). *Uchenye zapiski Petrozavodskogo gosudarstvennogo universiteta*, Nr. 1, pp. 42-46, 2014. [14-8-KS-Syn-sem-char-sep.pdf]

other contributions in proceedings:

[15-5] Hajičová Eva, Mikulová Marie, Panevová Jarmila: Reconstruction of Deletions in a Dependency-based Description of Czech: Selected Issues. In *Depling 2015: Proceedings of the Third International Conference on Dependency Linguistics*, Uppsala University, Uppsala, Sweden, pp. 131-140, 2015. [15-5-DepLing-elipsy.pdf]

[13-2] Kettnerová Václava, Lopatková Markéta: The Representation of Czech Light Verb Constructions in a Valency Lexicon. In *Depling 2013: Proceedings of the Second International Conference on Dependency Linguistics*, Matfyzpress, Praha, Czech Republic, pp. 147-156, 2013. [13-2-Depling-sep.pdf]

[14-9] Kettnerová Václava, Lopatková Markéta: Reflexive Verbs in a Valency Lexicon: The Case of Czech Reflexive Morphemes. In *EURALEX 2014: The User in Focus, Proceedings of the XVI EURALEX International Congress: EURAC research*, Bolzano/Bozen, Italy, pp. 1007-1023, 2014. [14-9-EURALEX-sep.pdf]

[15-6] Kettnerová Václava, Lopatková Markéta: At the Lexicon-Grammar Interface: The Case of Complex Predicates in the Functional Generative Description. In *Depling 2015: Proceedings of the Third International Conference on Dependency Linguistics*, Uppsala University, Uppsala, Sweden, pp. 191-200, 2015. [15-6-DepLing-LVC.pdf]

[12-4] Kettnerová Václava, Lopatková Markéta, Bejček Eduard: The Syntax-Semantics Interface of Czech Verbs in the Valency Lexicon. In *EURALEX 2012: Proceedings of the 15th EURALEX International Congress*, Department of Linguistics and Scandinavian Studies, University of Oslo, Oslo, Norway, pp. 434-443, 2012. [12-4-EURALEX-sep.pdf]

[13-3] Kettnerová Václava, Lopatková Markéta, Bejček Eduard, Vernerová Anna, Podobová Marie: Corpus Based Identification of Czech Light Verbs. In *Gajdošová, K., Žáková, A. (eds.) Slovo 2013: Natural Language Processing*,

Corpus Linguistics, E-learning. Proceedings of the Seventh International Conference, RAM-Verlag, Lüdenscheid, Germany, pp. 118-128, 2013. [13-3-Slovko-sep.pdf]  
[14-10] Skwarska Karolína: Rol' leksicheskogo napolnenia v konstrukcijach s otnosheniem semanticheskoy diatezy v cheshskom, ruskom i pol'skom jazykach. [The Role of Lexical Realization in the Constructions Related through Semantic Diathesis]. In Motoki Nomachi, Andrii Danylenko, Predrag Piper (eds.) Proceedings from the 36th Meeting of the Commission on the Grammatical Structure of the Slavic Languages of the International Committee of Slavists, pp. 347-359, 2014. [14-10-Skwarska-Rol-sep.pdf]  
[13-4] Vernerová Anna, Lopatková Markéta: Towards Automatic Detection of Applicable Diatheses. In ITAT 2013: Information Technologies - Applications and Theory (Proceedings), CreateSpace Independent Publishing Platform, Prešov, Slovakia, pp. 10-17, 2013. [13-4-ITAT-diateze-sep.pdf]

## PhD theses:

[15-7] Bejček Eduard: Automatické propojování lexikografických zdrojů a korpusových dat. Ph.D. thesis, Charles University in Prague, Faculty of Mathematics and Physics, Prague, Czech Republic, 227 pp., 2015. [15-7-EB-thesis.pdf]

[12-5] Kettnerová, Václava: Lexikálně-sémantické konverze ve valenčním slovníku. PhD thesis, Charles University, Prague, Czech Republic, 220 pp., 2012. [12-5-VK-thesis.pdf]

## Master thesis:

[13-5] Labbé, Grégoire: Traitement de la valence de verbes de mouvement slovènes sur la base de la valence de verbes tchèques. Mémoire de master 1. Institut National des Langues et Civilisations Orientales / Univerzita Karlova v Praze, 64 pp., 2013. [13-5-Gregoire-Labbe.pdf]

## data/software:

[12-6] Lopatková Markéta, Kettnerová Václava, Bejček Eduard, Skwarska Karolína, Žabokrtský Zdeněk: VALLEX 2.6. Data/software, ÚFAL MFF UK, <http://ufal.mff.cuni.cz/vallex/2.6/>, 2012.

[15-8] Lopatková Markéta, Kettnerová Václava, Bejček Eduard, Vernerová Anna, Žabokrtský Zdeněk: VALLEX 3.0 - Valenční slovník českých sloves. Data/software, Charles University in Prague, Faculty of Mathematics and Physics, <http://ufal.mff.cuni.cz/vallex/3.0/>, 2015.

## oral contributions

[15-9] Bejček Eduard: Propojování valenčních slovníků. Talk, Seminar of Formal Linguistics, Prague, Czech Republic, Oct 2015.

[14-11] Kettnerová Václava: Czech Light Verbs from a Lexicographic Point of View. Talk, Seminar of Formal Linguistics, Prague, Czech Republic, Jun 2014.

[14-12] Kettnerová Václava: Lexikografický popis komplexních predikátů v češtině. Talk, Výzkum slovesné valence ve slovanských zemích včera a dnes, Slovanský ústav AV ČR, v.v.i., Instytut Slawistyki UW Varšava, Prague Czech Republic, Dec 2014.

[12-7] Skwarska Karolína: Combination of different types of diatheses in Czech, Russian and Polish. Oral presentation, Session of the Commission on the Grammatical Structure of Slavic Languages of the International Committee of Slavists in Nitra, Oct. 2012.

[14-13] Skwarska Karolína: Specifické vyjádření účastníků situace u verb dicendi v češtině, ruštině a polštině. Oral presentation, Session of the Commission on the Grammatical Structure of Slavic Languages of the International Committee of Slavists in Ljubljana, Sept. 2014.

[15-10] Skwarska Karolína: Gde konchaetsja sintaksis i nachinaetsja semantika pri opisani svojstv slavjanskogo glagola. Oral presentation, Congress ICCEES Makuhari, Aug. 2015.

[15-11] Skwarska Karolína: Reciprocita v češtině, ruštině a polštině. Zpracování ve valenčním slovníku. Oral presentation, Session of the Commission on the Grammatical Structure of Slavic Languages of the International Committee of Slavists in Moscow, Sept. 2015.

## b) využití technického a přístrojového vybavení pořízeného z grantových prostředků:

MFF UK: The costs on computer equipment renewal (PC components) were partially covered by the grant budget in accordance with the original project proposal.

SI CAS: No computer equipment was covered by the grant.

## c) stručná zpráva o průběhu a výsledcích zahraničních cest:

The traveling costs related to the participation of the team members at specialized conferences and workshops were covered (sometimes partially) from the grant budget. An active participation (presentation of the results achieved within the project) was required.

## 2012:

- EURALEX 2012 (International Congress of the European Association for Lexicography), Oslo, Norway, Aug. 7-11, 2012 (Lopatková, Kettnerová); paper presentation [12-4] and scientific cooperation; [http://www.euralex.org/proceedings-toc/euralex\\_2012/](http://www.euralex.org/proceedings-toc/euralex_2012/)
- Session of the Commission on the Grammatical Structure of Slavic Languages of the International Committee of Slavists, Oct. 15-19, 2012, Nitra, Slovakia (Skwarska); participation on working session of the Commission, presentation [12-7];
- TSD 2012 (International Conference Text, Speech, Dialog), Brno, Sept. 3-7, 2012 (Kettnerová); paper presentation [12-3]; <http://www.tsdconference.org/tsd2012/>
- Summer school "Quantitative Methods and Corpora in Linguistic Research", University of Pavia, Pavia, Italy, Sept. 10-21, 2012 (Vernerová, partially supported); her newly gained knowledge of the methods of distributional analysis was applied in tasks 2.1 and 2.2; <http://studiumanistici.unipv.it/?pagina=p&id=481>
- study stay: Warsaw, Poland, Nov. 9-18, 2012 (Skwarska, partially supported); an examination of scientific books and papers, especially those focusing on syntactic properties of Polish verbs and their lexicographic description, which are not accessible in the Czech Republic.

## 2013:

- ITAT 2013 (Informačné Technológie – Aplikácia a Teória), Donovaly, Slovakia, Sept. 11-16, 2013 (Vernerová); paper presentation [13-4]; <http://itat.ics.upjs.sk/2013/pmwiki.php>
- Slovko 2013 (International Conference on NLP, Corpus Linguistics, E-Learning), Bratislava, Slovakia, Nov. 13-15, 2013 (Kettnerová); paper presentation [13-3]; <http://korpus.juls.savba.sk/~slovko/2013/>
- International Seminar: Voprosy ruskoj instoricheskoj grammatiki i slavjanovedenija: k 90-letiju so dnja smerti I. V. Jagića, Petrozavodsk, Russia, Sept. 19-20, 2013 (Skwarska); paper presentation [14-8];
- Università degli studi di Bergamo, Bergamo, Italy, Dec. 4-6, 2013 (Skwarska); invited talk for students 'Issledovanije valentnosti v slavjanskix jazykax. Novyje perspektivy.'

## 2014:

- Prešovská univerzita v Prešove, Prešov, Slovakia, Feb. 7-8, 2014 (Lopatková); presentation at the workshop organized by doc. Martina Ivanová and her research team on Czech and Slovak research on valency;
- LREC 2014 (International Language Resources and Evaluation Conference), Reykjavik, Island, May 26-31, 2014 (Bejček, Vernerová, partially supported); poster presentations [14-5], [14-7]; <http://lrec2014.lrec-conf.org/en/>
- EURALEX 2014 (International Congress of the European Association for Lexicography), Bolzano, Italy, July 15-19, 2014 (Lopatková, Kettnerová, partially supported); paper presentation [14-9]; <http://euralex2014.eurac.edu/en/Pages/default.aspx>
- Session of the Commission on the Grammatical Structure of the Slavic Languages of the International Committee of Slavists, Ljubljana, Slovenia, Sept. 3-7, 2014 (Skwarska); oral presentation [14-13].

## 2015:

- DepLing 2015 (International Conference on Dependency Linguistics), Uppsala, Sweden, Aug. 24-26 2015 (Lopatková, Kettnerová, partially covered); paper presentation [15-6]; <http://depling.org/depling2015/>
- ITAT 2015 (Informačné Technológie – Aplikácia a Teória); SloNLP 2015 (Slovenskočeský NLP workshop), Spišská Nová Ves, Slovakia, Sept. 17-21 2015 (Lopatková); consultations, member of the program committee; <http://itat.ics.upjs.sk/2015/pmwiki.php>
- ICCEES IX World Congress 2015: Makuhari, Japan, Aug. 3-8 2015 (Skwarska); oral presentation [15-10]; <http://www.l.u-tokyo.ac.jp/makuhari2015/index.html>
- Session of the Commission on the Grammatical Structure of Slavic Languages of the International Committee of Slavists, Moscow, Russia, Sept. 18-19 2015 (Skwarska); presentation [15-11].

## d) celkové shrnutí řešení projektu (u mezinárodních projektů včetně mezinárodní spolupráce):

The accomplishments of the four year project Delving Deeper: Lexicographic Description of Syntactic and Semantic Properties of Czech Verbs are twofold:

1. The theoretical insight into various language phenomena at the syntactic-semantic interface has been deepened – both grammaticalized alternations (diatheses and reciprocity) and lexicalized alternations (lexical-semantic conversions, structural splitting of a situational participant and multiple structural realization of a situational participant) were put under scrutiny. A contrastive perspective was also applied: we have focused namely on differences in syntactic behavior of Czech, Polish and Russian verbs undergoing changes in their valency structure. Moreover, a typologically new type of syntactic change – related to syntactic reflexivity in Czech – was identified and studied in detail, esp. changes in morphemic expressions of verbal complementations conditioned by the long and clitic variant of the reflexive pronoun.

Based on these theoretical achievements, the formal model of the lexicon has been refined and enriched – as a result, the final model provides an adequate and economic lexicographic representation of the studied phenomena. Further, elaborate lexicographic rules for describing changes in valency behavior of verbs undergoing alternations have been designed.

2. The electronic Valency lexicon of Czech verbs VALLEX has been substantially qualitatively and quantitatively enhanced in two dimensions. The first dimension concerns the data component. The information on the applicability of individual alternations (both grammaticalized and lexicalized) was added to its data component. Moreover, the lexicon has been enriched with the sample annotations of Polish and Russian lexical units undergoing alternations. Finally, VALLEX has been interlinked with the annotation lexicon of the Prague Dependency Treebank; as a result of the interlinking, the lexicon has been enriched with examples from the treebank.

The second dimension concerns the grammar component. Within the project, the VALLEX lexicon was enhanced with the grammar component specifying especially grammatical aspects of individual grammaticalized as well as lexicalized alternations.

The VALLEX lexicon is accessible on the following webpage <http://ufal.mff.cuni.cz/vallex/3.0/>.

The results were made available to the research community in journals dedicated to Czech as well as other Slavic languages (categories Jimp, Jneimp and Jrec; 7 articles in total), in thematic anthologies and as chapters of monographs (category C; 4 chapters), in 1 theoretical monograph (category B) and as the lexicon itself (category B). In addition, these results were presented at international and Czech conferences on both theoretical and computational linguistics (especially at those with proceedings monitored in WoS, category D; 5 published papers). Further, 2 more articles and 8 contributions in proceedings not registered in the Methodology were prepared and published.

The main applied output of the project is both qualitatively and quantitatively enhanced valency lexicon of Czech verbs available for a wide professional audience as well as for students and other language users. An emphasis was laid on both human and machine-readability; thus both linguists and developers of applications within the Natural Language Processing domain can use it.

The lexicon has been released as an electronic language resource (software, category R, open access).

Another positive aspect of the project was the involvement of students, which resulted in 1 MSc thesis (Gregoire Labbé) and 2 PhD theses (Václava Kettnerová and Eduard Bejček). Five more PhD students also participated in the project (Anna Vernerová, Marie Podobová, Natalia Klyueva, Katarzyna Vaculová and Adriana Filas) and thus have obtained important research experience.

e) přehled výsledků řešení grantového projektu za celou dobu řešení grantového projektu:

e1) Celkové počty výsledků:

1a. článek v odborném periodiku impaktovaném (druh výsledku Jimp)	1
1b. článek v odborném periodiku obsaženém v databázi Scopus (druh výsledku Jsc)	0
1c. článek v odborném periodiku neimpaktovaném (druh výsledku Jneimp)	3
1d. článek v českém odborném recenzovaném časopise (druh výsledku Jrec)	3
2a. odborná kniha (druh výsledku B)	2
2b. kapitola v odborné knize (druh výsledku C)	4
3. článek ve sborníku (druh výsledku D)	5
4. patent (druh výsledku D)	0

e2) Uplatněné výsledky řešení projektu v klasifikaci odpovídající platné Metodice hodnocení výsledků výzkumu a vývoje

**A1a. Jimp** (článek v odborném periodiku, který je obsažen v databázi Web of Science společností Thomson Reuters)

Autor, autoři: Kettnerová Václava, Lopatková Markéta, Panevová Jarmila

Název: Shoda doplňku v reflexivních konstrukcích v češtině.

Název časopisu, rok, volume, stránka od - do: Slovo a slovesnost, 2015, Vol. 76, No. 3, pp. 198-214

Dedikace na projekt: Ano

Vícenásobná dedikace:

**A1b. Jsc** (článek v odborném periodiku, který je obsažen ve světově uznávané databázi Scopus)

*Nic nezádáno*