

- 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 dílčí výsledky v daném kalendářním roce (u mezinárodních projektů včetně mezinárodní spolupráce):

In the third year of the project, all research areas were surveyed in accordance with the original proposal of the project. The accomplishments of the project are reflected in the following list of publications (the numbers refer to the table with project outputs): one book [3] and one chapter in book [4], and the journal articles [1] and [2]. Moreover, the achievements of the project were presented at four international conferences and workshops (focused on formal linguistics, natural language processing, and lexical resources) and published in their conference proceedings [6], [5], [8], and [7] as well as in the proceedings of the Meeting of the Commission on the Grammatical Structure of the Slavic Languages of the International Committee of Slavists (henceforth Grammatical Commission) [10]. In addition, four oral presentations were delivered at the conference devoted to the valency in Slavic languages [12], [13], [14], [16], one in the Seminar of Formal Linguistics [11], and one at the Meeting of the Grammatical Commission [15]. One journal article was submitted to a journal with impact factor [17] and two articles were submitted to reviewed journals [18], [19].

The research teams of the collaborating institutions – the Institute of Formal and Applied Linguistics (MFF UK) and the Slavonic Institute (SI CAS) – closely collaborated: an internal seminar was organized in June, attended by all team members as well as by other researchers; research plans, partial results as well as open problems and possible solutions were intensively discussed via e-mail. Members of both teams participated in the international valency conference held in Prague (Research of Verbal Valency in Slavic Countries in the Past and at Present), organized by the Slavonic Institute of the Czech Academy of Sciences and the Institute of Western and Southern Slavic Studies of the University of Warsaw.

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: In accordance with the proposed plan, theoretical research into changes in the valency structure of Czech verbs was carried on, and our attention shifted from central changes (esp. diatheses) to peripheral ones which have not been analyzed so far. These changes are associated with syntactic reflexivity. The adequacy of the lexicographic model of the lexicon (proposed in the previous years of the project) has been verified with respect to these peripheral changes. The proposed lexicographic representation – based on the division of the lexicon into the data and grammar components – proved to be as suitable for the description of these peripheral changes in the valency structure of verbs as for the central ones, see esp. [5]. Moreover, the necessity of a close cooperation between the grammar and the lexicon was exemplified on the description of the valency structure of Czech verbs, on grammatical diatheses and on light verbs presented in the oral presentations [14], [13], and [12], respectively.

SI CAS: The formal model of a trilingual valency lexicon was further refined. Some particular issues were discussed – variants of reflexive/non-reflexive verbs, stylistic equivalents, acceptability of Internet examples, formal realization of participants (sets of prepositions, conjunctions, participles etc.) [19].

1.2 Grammatical and syntactic diatheses: theoretical and practical aspects

MFF UK: The work on diatheses primarily focused on detailed syntactic rules for individual diatheses – the rules were further tested, refined, and evaluated with respect to the corpus data. Further, methods for the automatic identification of diatheses applicable for individual lexical units were improved and applied to the data; the automatic processing was followed by a manual annotation (see also sect. 1.6).

The theoretical model for capturing diatheses in the lexicon as well as the progress of the annotation were presented at two conferences [8], [13].

SI CAS: The annotation of Russian and Polish lexical units continued in accordance with the original plan.

A group of Polish verbs – those that enter structural splitting of a situational participant, see 1.3 – was selected for the annotation of grammatical diatheses (passive, deagentive, resultative). A specifically Polish reflexive construction with the accusative case has been annotated as well. We focused on reciprocity in Polish and Russian, especially on the expressions of reciprocity by either lexical or grammatical means.

Reflexive constructions in Polish in comparison with Czech were described in [16]. A detailed description of Czech resultative constructions was given in [18].

1.3 Semantic diatheses: theoretical and practical aspects

MFF UK: In the third year of the project, all three types of alternations – lexical-semantic conversions, structural splitting of a situational participant and multiple structural realization of a situational participant (in our previous terminology semantic diatheses) were described and richly documented by corpus evidence in the chapter “Alternace slovesných rámců” of the monograph “Mluvnice současné češtiny 2” [4]. Further, the lexical-semantic conversions in Czech were surveyed in the book “Lexikálně-sémantické konverze ve valenčním slovníku” [3].

In addition to theoretical research, the lexical rules describing changes in the correspondence between situational

participants and valency complementations of verbs were formulated and their list is prepared to be added into the grammar part of the lexicon.

The inventory of verbs allowing these types of alternations (compiled in the second year of the project) was projected into the data component of the lexicon.

SI CAS: In the comparative part of the project, primarily verbs entering the “subject – object” lexical semantic conversions were studied. Significant differences among Czech, Polish and Russian were detected: a lot of non-personal constructions expressing different types of meaning are used in Russian, unlike in Polish and esp. Czech. It is not possible to classify such constructions as typical alternations, and we have to formulate exact rules for annotation to achieve required consistency of their description in the valency lexicon. More detailed study is necessary in the future.

We also focused on Russian and Polish verbs with non-conversive lexicalized alternations, i.e. multiple structural expression of a participant and structural splitting of a situational participant. Structural splitting of a situational participant with verbs of perception and verbs of communication (verba dicendi) in Russian, Czech and Polish was described [15], [19].

A list of verbs with structural splitting of a situational participant in Polish was also compiled.

1.4 Other types of changes in the valency structure of verbs

MFF UK: This year, Czech reflexivity (in a broad sense) was studied in depth. We demonstrated that the Czech reflexive morphemes constitute different linguistic meanings (either they are formal means of the word formation process of the so called reflexivization, or they are associated with the syntactic reflexivity, reciprocity, and diatheses) and that they require to be lexicographically described in different ways, see [5]. Moreover, the last three mentioned linguistic phenomena bring about different changes in the valency structure of verbs. Esp. syntactic reflexivity was put under scrutiny in relation to changes in the valency frames of verbs: The issue whether the distinction between the long and clitic variant of the Czech reflexive pronoun results in the congruence changes of complements was widely debated in [1] and in the submitted journal paper [17].

Furthermore, Václava Kettnerová debated a lexicographic representation of Czech light verbs, linguistic phenomena closely associated with changes in the valency structure of verbs and nouns, in the oral presentations [11] and [12].

1.5 Comparative aspects of diatheses

MFF UK: In accordance with the original project description, only marginal research tasks were addressed at MFF UK. A case study of a group of Czech and Slovenian verbs with the prefix ‘vy-’ and ‘iz-’ was further elaborated. Further, an attempt to exploit the VALLEX lexicon for automatic derivation of valency characteristics of related languages (for the purposes of NLP) was reported [7].

SI CAS: The annotation of other groups of Russian and Polish verbs was carried out with the help of an electronic tool prepared at MFF UK (attributes: ‘synon’, ‘example’, ‘control’, ‘aspect’, ‘diat’, ‘reciprocity’). The evidence from the Russian National Corpus (www.ruscorpora.ru), Polish National Corpus (www.nkjp.pl) and from Internet texts was used and partly corrected by native speakers.

Comparative aspects of changes in valency structure studied from the theoretical point of view were described in Sections 1.2, and 1.3 above.

1.6 Application in an electronic lexical resource

MFF UK: The system for semiautomatic identification of lexical units of Czech verbs that may enter a particular type of diathesis was applied to the full VALLEX data for the recipient and possessive resultative diathesis, and to a sample of the data for the passive diathesis. The annotation has in large part confirmed the syntactic rules and requirements given by traditional grammars; instances of structures that are similar to diatheses but do not fully conform to the standard definitions have also been noted and will be further investigated. The annotation has also revealed a large number of instances of the alternation between DIR3 and LOC in the marked members of diatheses.

Czech verbs contained in the VALLEX lexicon exhibiting lexicalized alternations either of conversive or non-conversive character were documented (see Section 1.3) and their inventory was integrated in the new version of the lexicon (by means of the developed annotation tool).

The realization of the grammar component of the lexicon has been postponed until it is decided if and how to capture these phenomena (see also task 1.2).

As for the technological part of the project, new release of the VALLEX lexicon has been prepared:

- Source text data were cleaned and well-formed for better manipulation.
- External annotation of lexicalized alternation (Sect 1.3 and Sect 1.4) was merged with the original lexicon data.

As a result, already annotated grammatical as well as lexicalized alternations of Czech verbs were made public both as data in XML format and as a browsable new version of the web lexicon, VALLEX 2.6.3. <http://ufal.mff.cuni.cz/vallex/2.6.3/>.

SI CAS: The electronic annotation tool for Russian and Polish verbs was refined and implemented (linking of Czech-Russian-Polish counterparts, processing lemma variants in Polish etc.) in close cooperation with E. Bejček (MFF UK).

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

2.1 Enhancing Czech valency lexicon with semantic classes and semantic roles

Several experiments focusing on linking the VALLEX valency frames and the FrameNet semantic frames have been carried out. However, despite promising results, a detailed manual analysis has showed that the gained accuracy of the output is not satisfactory yet.

2.2 Strengthening the lexical resource with corpus evidence

The linking of the valency lexicons VALLEX and PDT-VALLEX has been further refined, evaluated and compared to manual annotations. As IAA (inter-annotator agreement) slightly exceeds 90%, the reported automatic results (precision 81) represent quite satisfactory output [6] and were projected to the data – the web-based version of the VALLEX lexicon contains the links to the PDT-VALLEX lexicon either on the level of individual lexical units (in cases where the reliability of the automatic linking reaches a given score) or on the level of whole lexemes (otherwise), see <http://ufal.mff.cuni.cz/vallex/2.6.3/>.

The automatic alignment between PDT-Vallex and Engvallex (Urešová et al, 2013) has not been used yet as the linking is not available yet (the alignment between PDT-Vallex and Engvallex is an external resource, which is not part of our project).

In addition to the work on the project areas, the SI CAS team organized a scientific conference entitled "Research of verbal valency in Slavic countries in the past and at present" in cooperation with the Institute of Western and Southern Slavic Studies, University of Warsaw. The conference took place in Prague, December 11–12, 2014. 28 researchers from the Czech Republic, Slovakia, Poland, Russia, and Croatia participated in the conference and they gave 22 talks on VALLEX, Walenty (valency lexicon of Polish verbs) and Valency dictionary of Croatian verbs, as well as on comparative research of syntactic and semantic structures of Slavonic verbs. Different approaches to verbal valency in current and former linguistics were presented. Selected papers will be published.

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

Bibliography

- [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
- [2] Podobová Marie: Dynamické a statické vyjádření "lokace" v češtině a v ruštině. Nová čeština doma a ve světě 1-2/2013, s. 94-109.
- [3] Kettnerová Václava: Lexikálně-sémantické konverze ve valenčním slovníku. Karolinum, Prague, Czech Republic, 200 pp., 2014
- [4] Kettnerová Václava, Lopatková Markéta: Alternace slovesných rámců. Chapter in: Panevová Jarmila, Hajičová Eva, Kettnerová Václava, Lopatková Markéta, Mikulová Marie, Ševčíková Magda: Mluvnice současné češtiny 2, Syntax na základě anotovaného korpusu. Karolinum Praha, Prague, Czech Republic, 291 pp., 2014
- [5] Kettnerová Václava, Lopatková Markéta: Reflexive Verbs in a Valency Lexicon: The Case of Czech Reflexive Morphemes. In: Proceedings of the XVI EURALEX International Congress: The User in Focus, EURAC research, Bolzano/Bozen, Italy, pp. 1007-1023, 2014
- [6] Bejček Eduard, Kettnerová Václava, Lopatková Markéta: Automatic Mapping Lexical Resources: A Lexical Unit as the Keystone. In: Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014), European Language Resources Association, Reykjavík, Iceland, pp. 2826-2832, 2014
- [7] Klyueva Natalia, Kuboň Vladislav: Automatic Valency Derivation for Related Languages. In: Proceedings of the Twenty-Seventh International Florida Artificial Intelligence Research Society Conference, FLAIRS 2014, Press, Palo Alto, California, pp. 437-442, 2014
- [8] Vernerová Anna, Kettnerová Václava, Lopatková Markéta: To pay or to get paid: Enriching a Valency Lexicon with Diatheses. In: Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014), European Language Resources Association, Reykjavík, Iceland, pp. 2452-2459, 2014.
- [9] Skwarska Karolína: Sintaksicheskiye i semanticheskiye charakteristiki russkix, pol'skix i cheshskix glagolov v slovare sochetajemostej VALLEX. (Syntactic and semantic properties of Russian, Polish and Czech verbs in valency lexicon VALLEX). Uchenyye zapiski Petrozavodskogo gosudarstvennogo universiteta 2014, Nr. 1, s. 42-46.
- [10] Skwarska Karolína: Rol' leksicheskogo napolnenia v konstrukcijach s otnosheniem semanticheskoy diatezy v cheshskom, russkom i pol'skom jazykax. [The Role of Lexical Realization in the Constructions Related through Semantic Diathesis]. Proceedings from the 36th Meeting of the Commission on the Grammatical Structure of the Slavic Languages of the International Committee of Slavists. Edited by Motoki Nomachi, Andrii Danylenko, Predrag Piper. 2014.
- [11] Kettnerová Václava: Czech Light Verbs from a Lexicographic Point of View. Talk, Seminar of Formal Linguistics,

Prague, Czech Republic, Jun 2014

[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 Zachodniej i Południowej UW Varšava, Prague Czech Republic, Dec 2014

[13] Lopatková Markéta, Vernerová Anna: Zaplatit nebo dostat zapláceno: Zachycení diatezí ve Valenčním slovníku českých sloves VALLEX. Talk, Výzkum slovesné valence ve slovanských zemích včera a dnes, Slovanský ústav AV ČR, v.v.i., Instytut Slawistyki Zachodniej i Południowej UW Varšava, Prague Czech Republic, Dec 2014

[14] Panevová Jarmila: Valence v gramatice, valence ve slovníku. Talk, Výzkum slovesné valence ve slovanských zemích včera a dnes, Slovanský ústav AV ČR, v.v.i., Instytut Slawistyki Zachodniej i Południowej UW Varšava, Prague Czech Republic, Dec 2014

[15] Skwarska Karolína: Specifické vyjádření účastníků situace u verb dicendi v češtině, ruštině a polštině. Zasedání Komise pro gramatickou strukturu slovanských jazyků, Lublaň 3.-5.9.2014. Oral presentation.

[16] Skwarska Karolína, Bolbot Katarzyna, Filas Adriana: Zpracování polských zvrtných konstrukcí ve slovníku VALLEX. Talk, Výzkum slovesné valence ve slovanských zemích včera a dnes, Slovanský ústav AV ČR, v.v.i., Instytut Slawistyki Zachodniej i Południowej UW Varšava, Prague Czech Republic, Dec 2014.

[17] Kettnerová Václava, Lopatková Markéta, Panevová Jarmila: Shoda doplňku v reflexivních konstrukcích. In: Slovo a slovesnost, pp. 1-19 (submitted)

[18] Podobová Marie: Bezobjektové posesivní rezultativní konstrukce od sloves tranzitivních a bezobjektové i objektové posesivní rezultativní konstrukce od sloves intransitivních - jejich tvorba a funkce v českém jazyce. Nová čeština doma a ve světě 1/2015 (accepted)

[19] Skwarska Karolína: Opis leksykograficzny syntaktycznych i semantycznych cech czasowników w języku polskim i rosyjskim na podstawie korpusów narodowych. Prace Filologiczne. (submitted)

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 budget in 2014.

c) stručná zpráva o průběhu a výsledcích zahraničních cest:

- Prešovská univerzita v Prešově (Prešov, Slovakia, Feb. 7-8, 2014); Markéta Lopatková, presentation at the workshop organized by doc. Martina Ivanová and her research team on Czech and Slovak research on valency;
- LREC (International Language Resources and Evaluation Conference, Reykjavik, Island, May 26-31, 2014); Anna Vernerová, poster presentation [8]; Eduard Bejček, poster presentation [6];
- EURALEX conference (Bolzano, Italy, July 15-19, 2014); Markéta Lopatková, Václava Kettnerová, paper presentation [5];
- Meeting of the Commission on the Grammatical Structure of the Slavic Languages of the International Committee of Slavists (Ljubljana, Slovenia, Sept. 3-7, 2014); Karolína Skwarska, oral presentation [15].

d) plán práce na další rok řešení (u mezinárodních projektů včetně mezinárodní spolupráce):

During the last year of the project, research into the identified areas in accordance with the original proposal will be conducted, with slight modifications following from the already achieved goals. Both theoretical and practical aspects will be addressed, with the aim to project all results onto the data of the valency lexicon of Czech verbs, VALLEX. An important part of the working capacity will be used for publishing results at specialized conferences and in journals. The work on the planned monograph (printed version of the lexicon) will be finished – the introductory study will be composed, the lexicon data will be transferred to the printable version and a book layout of the valency entry will be designed. The lexicon data will be released in the form of a web-based lexicon.

1 The lexicographic representation of various changes in valency structure of verbs

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

Based on final annotations, potential minor aspects of the formal model of lexicographic representation of Czech verbs will be settled, and they will be projected to the lexicon data (as described in the following sections).

1.2 Grammatical and syntactic diatheses: theoretical and practical aspects

MFF UK: The annotation of diatheses (see tasks in sect. 1.6) has mostly confirmed the syntactic rules and requirements given by traditional grammars. However, instances of structures that are similar to diatheses but do not fully conform to the standard definitions have also been noted and will be further investigated. The annotation has also revealed a large number of instances of the alternation between DIR3 and LOC in the marked members of diatheses. The proposed structure of the grammar component of the lexicon will be revisited to capture these phenomena in the lexicon.

SI CAS: The theoretical comparative research on grammatical diatheses (Czech-Russian-Polish) will continue, especially on the Polish and Russian resultative constructions and reciprocity.

1.3 Semantic diatheses: theoretical and practical aspects

MFF UK: The inventory of lexicalized alternations integrated in the lexicon will be checked with respect to its completeness and adequacy and (if necessary) it will be further modified. Moreover, a close cooperation of the lexicalized alternations and the list of the lexical rules describing lexicalized alternations formulated in the previous year of the project will be verified and further refined (if necessary). In accordance with our theoretical findings, the lexicographic representation of the lexicalized alternations will be completed in the VALLEX lexicon (see 1.6).

SI CAS: The annotation of Russian and Polish lexical units – counterparts of Czech lexical units of verbs entering lexicalized alternations in VALLEX 2.6.3 – will be finished. We will focus especially on the lexical-semantic 'subject – object' conversions in Russian and Polish both from the theoretical and from the practical point of view.

1.4 Other types of changes in the valency structure of verbs

MFF UK: Syntactic reflexivity will be further theoretically studied in the last year of the project and its lexicographic representation will be proposed in order to cover also peripheral changes in the valency structure of Czech verbs. The main focus will be laid on the distinctions between the clitic and long variant of the reflexive pronoun and its impact on the surface structure of a sentence.

SI CAS: The annotation of relevant Polish and Russian lexical units will be finished.

1.5 Comparative aspects of diatheses

MFF UK and (mainly) SI CAS: All the above mentioned types of alternations (diatheses, reciprocity, lexicalized alternations) will be discussed.

1.6 Application in an electronic language resource

MFF UK: The inventory of lexicalized alternations compiled in 2013 and integrated in the VALLEX lexicon in 2014 will be tested with respect to its completeness and adequacy. Moreover, lexical rules reflecting changes in correspondence between situational participants and valency complementations will be further refined (if necessary) and their inventory will be verified on the linguistic data already projected into the VALLEX lexicon. The main goal is the data and the grammar component of the lexicon enhanced with the information on lexicalized alternations.

The final version of the VALLEX lexicon will be released. It will contain information about diatheses (1.2) and alternations (1.3 and 1.4), semantic classes (2.1) and corpus evidence (2.2). The main output will be the printed version of VALLEX, which will be prepared and typeset by the research group and published by Karolinum publishing house. In addition, a new web interface for the web-based lexicon will be developed and the lexicon will be released.

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

2.1 Enhancing Czech valency lexicon with semantic classes and semantic roles

The effort of linking the VALLEX lexicon with FrameNet will continue, with the aim to enrich VALLEX with missing

semantic information. Both Semantic Frames (semantic classes) and Frame Elements (semantic roles) will be connected to the appropriate equivalent in VALLEX, the semantic relation of 'Inheritance' encoded in FrameNet will be used. Moreover, this mapped semantic information will be added to the HTML version of the lexicon.

2.2 Strengthening the lexical resource with corpus evidence

MFF UK: At the beginning of the year, VALLEX will be enriched with corpus evidence from the Prague Dependency Treebank. Several sentences (either from PDT or from CNK obtained in VALEVAL project) for each lexical unit will be selected for the printed version of the lexicon.

e) souhrn výsledků dosavadního řešení projektu a naplnění cílů projektu:

Within the first three years of the project *Delving Deeper: Lexicographic Description of Syntactic and Semantic Properties of Czech Verbs*, we have already achieved significant results relevant to the general goals of the project.

1. The theoretical insight into various phenomena at the syntax-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. The 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 reflexivization – 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 were analyzed.

Based on these 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. The elaborated lexicographic rules for describing changes in valency behavior of verbs undergoing alternations have been proposed; after their evaluation in the last year of the project, their form will be fixed.

2. The electronic Valency lexicon of Czech verbs VALLEX has been substantially qualitatively and quantitatively enhanced in two dimensions. First, 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 verbs undergoing alternations (not publicly available yet). Second, 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 the examples from the treebank.

<http://ufal.mff.cuni.cz/vallex/2.6.3/data/html/generated/alphabet/index.html>

The results were made available to the research community in 7 journal articles, 9 papers in conference proceedings, 1 monograph, 1 book chapter and several oral presentations; manuscripts of further journal submissions have also been prepared.

Another positive aspect of the project is the involvement of students, which resulted in 1 MSc (Gregoire Labbé) and 1 PhD thesis (Václava Kettnerová). Six more PhD students also participate in the project (Eduard Bejček, Anna Vernerová, Marie Podobová, Natalia Klyueva, Katarzyna Vaculová and Adriana Filas) and thus they obtain important research experience.