Verb Valency and Argument Non-correspondence in a Bilingual Treebank

Jana Šindlerová, Zdeňka Urešová, and Eva Fučíková

Charles University in Prague
Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics
Czech Republic

Abstract. In this paper we present a contrastive study of one interesting non-correspondence between deep syntactic valency structures of two different languages. On the material of the Prague Czech-English Dependency Treebank we observe sentences in which an Addressee argument in one language is linked translationally to a Patient argument in the other one, particularly we aim our attention at the class of judgement verbs (in a broad sense). Considering this class of verbs, we analyze the relevant examples and discuss the nature of "the third argument" in the valency structure. As a result, we reconsider the conventions of argument labelling with the aim of achieving better consistency of annotation and we suggest possible ways of adjusting the valency theory itself to the needs of multilingual data.

1 Introduction

Modern approaches to applied linguistics take the advantage of a great number of annotated corpora, covering different depth and width of linguistic description, a wide range of content domains, and above all an impressive scale of world languages. Many of these corpora are accompanied by additional resources, such as valency lexicons. Parallel valency lexicons, accompanying multilingual corpora, satisfy the call for capturing complex lexical information, i.e. the information on both verbal translational equivalents and their valency slot realizations. Having resources of this kind at our disposal gives us a perfect opportunity to study similarities and differences between languages on the syntax-semantics interface.

In this paper we will focus on Czech and English deep syntactic valency structures in a contrastive perspective. The assumption we follow is that the deeper we look into the linguistic structure, the more similar should the structures appear, an idea shared e.g. by [5]. This idea stands behind numerous attempts to create an interlingua for machine translation systems from various types of deep syntactic structures, or semantic representations.

Our goal within the research is to look at the points of non-correspondence between deep syntactic structures of Czech and English parallel sentences, to analyze them and categorize according to the syntactic and semantic properties of the utterances they represent. Questions we would like to ask in this paper are the following:

- Are there any semantic (or syntactico-semantic) criteria, rather than mere morphological hints, to let us distinguish clearly between possible variants of argument labelling?
- Does the cross-linguistic perspective offer a better insight into the nature of differences between the individual frames?
- Does the cross-linguistic perspective help us in deciding about possible theoretical amendments and in making the annotation practice more uniform?

2 Methodology and Data

We took the advantage of the existence of Czech-English parallel data, namely the Prague Czech-English Dependency Treebank (PCEDT) [4]. It is a collection of about 50 000 sentences, taken from Penn Treebank-Wall Street Journal section, translated manually to Czech, transformed into dependency trees and annotated at the level of deep syntactic relations (called the tectogrammatical layer). In short, the tectogrammatical layer contains mostly content words (with several defined exceptions) connected with oriented edges and labelled with syntactico-semantic functors according to the Functional Generative Description approach (FGD), see [6]. Ellipsis and anaphora resolution is also included, as well as an automatic alignment of corresponding nodes [12].

The PCEDT 2.0 [3] is annotated according to the FGD valency theory and two valency lexicons (one for each language) are part of the release of the treebank. The PDT-Vallex [15, 16] has been developed as a resource for annotating argument relations in the Prague Dependency Treebank [1]. Valency frames in the PDT-Vallex roughly correspond to individual verb meanings. Valency frames consist of participant slots represented by tectogrammatical functors. Each slot is marked as obligatory or optional and its typical morphological realization forms are listed. Frame entries are supplemented with illustrative sentence examples.

The Engvallex was created as an adaptation of an already existing resource of English verb argument structure characteristics, the Propbank. The original Propbank argument structure frames have been adapted to the FGD scheme, so that it currently bears the structure of the PDT-Vallex, though some minor deflections from the original scheme have been allowed in order to save some important theoretical features of the original Propbank annotation.

Currently, there is a project aimed at interlinking PDT-Vallex and Engvallex in the sense of gaining a database of frame-to-frame, and subsequently, slot-to-slot pairs for the purposes of machine translation experiments [17], extending a similar project held in the past [14].

In the project we also deal with semantic categories and verb classes. Since this topic is not covered within the FGD theory, we have consulted other available resources of native speaker's intuition regarding valency characteristic of English verbs: the Propbank [9], the Framenet [13] and Levin's classification (as stated in [10]).

3 Argument Labelling in PDT 2.0

In the FGD, five *actants*, i.e. main universal and typical arguments of a valency frame, are recognized: ACT (Actor), PAT (Patient), ADDR (Addressee), EFF (Effect) and ORIG (Origin). In the theoretical framework, it is declared that ACT and PAT stand for more general concepts of "the first" and "the second" argument in the valency structure, in other words, these positions are described more syntactically than semantically. On this theoretical background, the concept of "shifting of cognitive roles" has been adopted. According to this rule, if the verb has only two arguments, semantic Effect, semantic Addressee and/or semantic Origin are being shifted to the Patient position. With respect to such definition, we may encounter several difficulties in our research. Typically, if we look for the differences in argument labelling, we may easily be confused by different labelling of (semantically the same) Addressee argument due to a different number of valency positions in the corresponding frames, as in (1).

(1a) John.ACT blamed Mary.ADDR of stealing.PAT his car.
(1b) John.ACT shouted at Mary.PAT.

Another problem is tied with cases of three (and more) recognizable participants in the valency structure, where ADDR in fact appears as the (syntactically) second argument (often expressed with accusative), whereas the PAT label is left for another argument role. Though much attention is paid to the criteria for the theoretical distinction of actant and free modifier roles, and for the distinction of obligatory and optional positions in FGD, surprisingly little is said about the nature of individual actant roles per se. It is somehow taken for granted that native speaker intuition in this respect recognizes semantic aspects of the actant roles well. For example, PDT guidelines [12] describe PAT as an "affected object" in a broad sense and offer an illustrative (non-exhaustive) list of its possible semantic modifications, but only for PAT as a second argument, leaving out the (for our research) interesting cases where PAT takes a third position in the valency structure[1], syntactically realized as a prepositional phrase following a direct object ADDR, like in (1a).

The authors of the PDT 2.0 annotation guidelines also confess there is a certain degree of uncertainty about the character of arguments with certain verbs and explicitly mention several borderline cases [12]. As a typical case, they offer an example of the Czech verb *bránit (protect)*, with the following possible interpretations of the three available arguments (Protector, Protected, Harm/Enemy):

(2a) Petr.ACT bránil majetek.PAT před zloději.EFF.

[1] Our numbering of argument positions in the paper is given by purely syntactic properties, i.e. subject first, (direct) object(s) second (or second and third), prepositional phrases third (fourth) etc. In English, this also corresponds to a standard word order; however, we also use it for Czech.

Petr.ACT protected his property.PAT from thieves.EFF.
(2b) Petr.ACT bránil dětí.ADDR před nebezpečím.PAT.
Petr.ACT protected the children.ADDR from danger.PAT

The resolution of this problem in the annotation guidelines is based on a morphosemantic feature of animacy. If the defended entity is represented in a majority of corpus occurrences by an animate noun (or, more precisely, by an animate entity), the position in the valency lexicon should be labelled Addressee, otherwise it is assigned a Patient label. Still, it is a common phenomenon that a degree of interannotator disagreement is noticed when dealing with similar cases.[2]

The investigation of a contrastive language material shows that such cases are frequent and tightly connected to the semantic class of the verb. Moreover, it appears that native speaker intuition differs in the contrastive point of view.

4 Places of Non-correspondence

When searching the PCEDT, we have encountered five major verb classes that show inconsistencies in the annotation of valency structures, mainly concerning the Addressee role: verbs of judgement, verbs of attempting suasion (and causation), a joint group of several classes semantically expressing permission or accessibility granting, verbs of assistance and verbs of commercial exchange.[3] For each of the verb classes, we have consulted several resources of valency structure description. Apart from the PDT-Vallex and the Engvallex, we have searched the Framenet and the Propbank. For the reference to semantic classes of verbs, we have also consulted Levin's classification. Throughout the five mentioned semantic groups, we have encountered several patterns of frame concurrence, the most frequent being the following:

- ACT ADDR PAT x ACT PAT CAUS
- ACT ADDR PAT x ACT PAT EFF or ACT PAT AIM
- ACT ADDR PAT x ACT PAT REG
- ACT ADDR PAT x ACT PAT MEANS

This concurrence of frames appeared both in the cross-linguistic comparison (a source language sentence is annotated differently than the target language sentence) and within different verbs of the same semantic group in one language (two verbs of one language, which are semantically close, or even synonymous, are annotated differently).

All the above mentioned alternative frame variants consist of an all-actant interpretation (ACT ADDR PAT) and an interpretation involving an adjunct

[2] It is also interesting that other researches working within the FGD framework do not operate with animacy in this respect, cf. [11].

[3] The naming of the classes has been roughly adopted from the Framenet nomenclature.

on the third syntactic position (ACT PAT CAUS). For the purposes of simplification we do not operate here with the notion of obligatoriness or optionality, leaving this complex issue to a separate, more elaborate study. Nevertheless, it must be said that this issue is of supreme interest to us, since it is tightly connected to the elementary question of argumenthood. Although FGD theoretically allows free modification functors to appear as a part of valency frame in case they are obligatory, in reality, this accounts only for a few members of the list of possible free modifications, usually for directionals, temporals and manner adjuncts. On the other hand, it is not a common practice e. g. to label an obligatory argument CAUS, even if its semantic incorporates causal interpretation, in this case, usually an actant label is given priority.

4.1 Judgement Verbs in Czengvallex

There are two ways in which argument non-correspondence in a bilingual corpus can be considered. Either there is a different argument labelling between a particular sentence and its translation, or there is a difference in argument labelling between verbs of the same verb class within a particular language. Both types of argument non-correspondence manifest in PCEDT among the verbs of judgement and communicating judgment. In Framenet, these two categories are considered separate, for our purposes it seems convenient to treat them jointly, e.g. as they appear in Levin's classification [10]. In the analysis, we will refer to them as *judgement verbs*.

In our sample, we have looked at the following verbs: *accuse, blame, charge, chastise, convict, criticize, fault, reprove, sue*. According to the three resources of English verbs argument structures, these verbs share three argument roles, which can be characterized as follows: *the judge*, *the judged entity* and *the reason for judgment*, or *the fault*. In the PCEDT (and its valency lexicons), the annotation practice is divided as shown in Table 1. The individual rows of the table represent different translation verb pairs, columns show the distribution of the verbs among different frames. If both verbs of the translation pair belong to the same frame, they are both inscribed in the same cell.

Most of the verbs fall into one of the following frame variants: ACT ADDR PAT, or ACT PAT CAUS (the labels of the positions marking *the judge*, *the judged entity* and *the fault* respectively). All the mentioned verbs have an addressee (*the judged entity*), though in the other variant of the frame, it is labelled PAT due to the concept of shifting cognitive roles (see [12]). The split of the annotation is apparently caused by different approaches to the third argument, i. e. *the reason for judgement*, or *the fault*. Either it is interpreted as an actant, i.e. belonging to the valency structure, or it is considered an adjunct, a free modification external to the valency structure.

Clearly, the important question is, whether *the reason for judgement* position is or is not a part of the valency structure of the verb. The resources for En-

Table 1: Frame distribution for Czech and English Judgement Verbs in the PCEDT

ACT ADDR PAT	**ACT PAT CAUS**	**ACT PAT EFF CAUS**
obvinit – accuse		
obviňovat – accuse		
vinit – accuse		
charge (with) – obvinit		
charge (with)	obžalovat	
charge (with)	žalovat	
obvinit	blame (for)	
připisovat – blame (on)		
připisovat	blame (for)	
přisuzovat – blame (on)		
přičítat – blame (on)		
obvinit – convict		
usvědčit – convict		
convict		odsoudit
obvinit	fault	
reprove	odsuzovat	
sue	žalovat	
soudit se – sue		
vytknout	chastise	
	kárat – chastise	
	potrestat – chastise	
	kritizovat – criticize	

glish verbs speak straight, all regard this position as a valency argument.[4] This question is closely connected to the question of "what exactly is an argument (theta role, participant etc.) and how many of them there really are", which has not been satisfyingly answered in the literature yet. Since this type of problem is too complex to be dealt with within this paper, we will not try to answer it directly. A very nice and summarizing debate of this issue can be found e.g. in [2]. Not only does the author question the mere possibility of finding clear matching criteria for argumenthood that would apply to all types of arguments, but he also mentions an important catch of the argumenthood-defining efforts. When trying to describe argument roles unambiguously, we necessarily use criteria from many levels of linguistic description, not only syntactic and semantic, but we also have to engage morphological and pragmatical hints. This on one hand helps us to specify the roles more exactly, but on the other hand leads many times to confusion and theoretical clashes.

[4] For example in the Framenet, Reason is always part of the core frame elements of the judgement frames, Propbank also includes it in the list of frame participants of the verbs in question.

4.2 Data Analysis

As we can see from Table 1, the annotation is to a certain extent inconsistent. There are at least four different English verbs correctly translated by Czech verb *obvinit/obviňovat*.[5] In case of *accuse* and *charge*, the third argument is PAT, whereas in case of *blame* and *fault*, it is labelled CAUS, see (3).

(3a) Industrial companies.ACT are accusing financial institutions.ADDR of jeopardizing.PAT Japan's economy.
(3b) A Campeau shareholder.ACT charged Campeau.ADDR with violating securities law.PAT.
(3c) Many investors.ACT blamed program trading.PAT for aggravating.CAUS market swings.
(3d) The former New York City mayor.ACT faults Obama.PAT for incompetence.CAUS over the Libya consulate attacks.

Since there is no significant difference in the verb semantics, this may be the result of the influence of morphological form: The for-phrase (*blame, fault*) is a typical morphological means for expressing Cause, whereas the of-phrase (*accuse*) is one typical way of expressing Patient (affected object) semantics, and the with-phrase (*charge*) is typical for Instrument interpretation, thus being tentative to less specific labelling.

Speaking about the impact of morphosyntactic form, we must point out another fact. Whereas the direct object form of *the reason for judgement* argument builds almost immediately the actant interpretation, the prepositional phrases are ambiguous with respect to possible interpretations. According to the Prague annotation style, it appears that only primary prepositions are allowed with arguments, whereas phrases with secondary prepositions are generally regarded as adjuncts. For each actant and adjunct label, the guidelines offer a list of typical prepositional phrases (in the form of preposition plus case) used with it. The *reason for judgement* can be, with some, but apparently not all the verbs, expressed in a typically adjunct morphosyntactic form, e.g. with a subordinate adjunct clause or a secondary preposition.

(4) Vyšetřovatel obvinil kvůli incidentu u mosteckého klubu Neprakta tři muže.
The investigating officer.ACT charged three men.PAT because of the incident.CAUS in the Neprakta club in Most.[6]

Nevertheless, it seems to us that such utterances are actually less acceptable since they mix the intended "objective reason for judgement" semantic interpretation with the typical form of expressing a "circumstantial motivation for judgement" (like in case of (5)):

[5] In case of convict the translation may be considered inappropriate.
[6] Since there was no suitable example in the PCEDT data, the sentence has been taken from Czech National Corpus. The labelling has been added by the authors of the paper.

(5) Obvinila ho, protože zrovna neměla dobrou náladu.
She blamed him because she wasn't in a good temper at the time.

In such cases, the form influences the interpretation to the circumstantial one, and the reason for judgement appears as non-overt.

Note that considering the third argument itself, there are equally relevant reasons for both interpretations (Patient and Cause). The semantics of the argument in question bears causal features (Framenet e.g. names this role *Reason*). On the other hand, it is often expressed (in lexicalized alternations[7] of the verbs in question) in a direct object position, which is typical for Patient and atypical for Cause (6).

(6) Nobody.ACT would blame the global warming.PAT on a few hundred thousand hunter-gatherers.ADDR hunting mammoths and scratching around in caves.

What is even more confusing, not even the criterion of obligatoriness can support our decision between argument and adjunct interpretation, since in the PDT-Vallex, the PAT argument of the judgement verbs is often marked optional.

With the verb pair *odsoudit – convict*, the situation is even more interesting.

(7) Despite the strong evidence against Mrs. Yeargin, popular sentiment was so strong in her favor, Mrs. Ward says , that "I'm afraid a jury wouldn't have convicted her."
I přes přesvědčivé důkazy proti Yearginové bylo této učitelce veřejné mínění tak silně nakloněno, že ředitelka Wardová říká: "Obávám se, že by ji porota neodsoudila.

The Czech verb, according to the PDT-Vallex opens valency positions for *the judge* (ACT), *the judged entity* (PAT), and *the sentence* (EFF), *the reason for judgement* being considered an adjunct CAUS. On the other hand, the original English verb *convict*, as far as Framenet and Propbank annotation states, does not include *the sentence* in its argument structure at all. Once again, we may ask the question, what are the criteria for considering *the reason for judgement* an adjunct in Czech, and not (an optional) PAT, while *the judged entity* slot could be easily re-interpreted as ADDR.

Another theoretical question considering the number of valency positions is effected by the lexicalized alternations.[8] In both Propbank and Framenet, the criticized entity and the cause of critique with verbs of communicating judgement are distinguished and treated separately, so that cases like (8a) and (8b) (where any other overt realization of a for-phrase argument is unlikely) get two different frames, thus saving the syntactic and semantic difference.

[7] See [7]
[8] For more information on lexicalized alternations, see [8].

(8a) He.ACT criticized him.PAT for coming.CAUS late.
(8b) He.ACT criticized his coming.PAT late.

On the other hand, in the PDT-Vallex, these frames are often unified into a single one with an optional Cause argument, disregarding the fact that in the second case, it is hardly imaginable that another Cause argument, with the meaning of an *objective reason for judgement* should overtly appear.

5 Proposal

What we propose is that the labelling of arguments should be as uniform as possible[9] within semantically related verbs. In case of judgement verbs we find two possible variants of labelling for unification of the annotation practice.

The all-actant variant consists of ACT (*the judge*), ADDR (*the judged entity*), PAT (*the reason for judgement*), and eventually EFF (*the sentence*). Its advantages and disadvantages are listed below:

+ In the available resources for valency characteristics of the English verbs, *the reason for judgement* is considered a part of the inner argument structure of a judgement verb, disregarding its actual morphosyntactic form, the all-actant solution keeps it a part of the frame even if it is not obligatory. Since there may be different intuitions considering obligatoriness across languages, this is an advantage with respect to the task of collecting argument alignment between languages.
+ Our proposal enables us to treat uniformly all judgement verbs having both *the judged entity* and *the reason for judgement* in their argument structure. As a result, the tectogrammatical structures of parallel trees of different languages would appear more similar.
+ Such labelling enables us to treat uniformly lexicalized alternations of the type shown in (6) for individual verbs.
+ It also enables us to distinguish between the *reason for judgement* (PAT), which according to our opinion belongs into the valency frame of the judgement verbs, and the *circumstantial cause* (CAUS) which is an adjunct describing some less relevant circumstances of the situation.
− Since PAT is a semantically underspecified label, the semantics of *reason for judgement* is lost in the description.

The adjunct variant, including ACT (*the judge*), PAT (*the judged entity*), CAUS (*the reason for judgement*), and eventually EFF (*the sentence*), on the other hand, has the following implications:

+ The semantics of *the reason for judgement* stays explicit in the annotation.

[9] We are aware of the fact that there are certain frame alternations that could not be unified with respect to argument labelling in the current framework, still we see a relatively large number of inconsistencies that can be repaired by clarifying the vague points in the theory.

- *The reason for judgement* will often be left out of the frame, or the theory must be revised in order to allow obligatory (and maybe even optional) adjuncts of the CAUS type into the frame.
- It will not be possible to maintain uniform approach to verbs of the same verb class, since *the reason for judgement* gets into the position of an affected object with some of the verbs. In a cross-linguistic comparison, this will result in mismatches and unnecessary confusion.
- It will not be possible to distinguish clearly between objective and circumstantial cause of judgement.

It seems that the all-actant variant is in many respects more advantageous than the adjunct variant. Still, to make an ultimate decision, it would be necessary to make the description of argumenthood, and maybe even obligatoriness, more clear and deciding. Also, the analysis of other argument mismatches in other verb classes may help to get a more complex picture of this issue.

We would like to pin-point that our proposal does not aim at being universal or exhaustive. There may appear exceptional cases which do not fit into our description. We base our description on the assumption that judgement verbs from a class with to a great extent uniform morphosyntactic behaviour, which of course may not be the case of other verb classes.

6 Conclusion and Future Work

In this paper we have presented a cross-linguistic analysis of valency frame mismatches within one semantic class of verbs in a parallel corpus and its valency lexicons. On the example of judgement verbs, we have pointed out several weak points of the annotation rules and suggested that clarification and further specification of the theory should help in keeping the data more consistent. As an example, we have proposed a concrete way of unifying the annotation practice for the class of judgement verbs.

In the future, we would like to continue with the analysis of typical argument and frame mismatches for other argument pairs and verb classes, in order to gain a better insight into the conceptual character of argumenthood and obligatoriness.

7 Acknowledgements

The project has been partially supported by the grant No. GPP406/13/03351P of the Grant Agency of the Czech Republic and by the SVV project number 267 314.

This work has been using language resources developed and/or stored and/or distributed by the LINDAT-Clarin project of the Ministry of Education of the Czech Republic (project LM2010013).

Bibliography

[1] Böhmová, A., Hajič, J., Hajičová, E., and Hladká, B. (2003). The Prague Dependency Treebank. In *Treebanks*, pages 103–127. Springer.

[2] Dowty, D. (1991). Thematic Proto-Roles and Argument Selection. *Language*, 67(3):547–619.

[3] Hajič, J., Hajičová, E., Panevová, J., Sgall, P., Bojar, O., Cinková, S., Fučíková, E., Mikulová, M., Pajas, P., Popelka, J., Semecký, J., Šindlerová, J., Štěpánek, J., Toman, J., Urešová, Z., and Žabokrtský, Z. (2012). Announcing Prague Czech-English Dependency Treebank 2.0. In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)*, pages 3153–3160, Istanbul, Turkey. European Language Resources Association.

[4] Hajič, J., Hajičová, E., Panevová, J., Sgall, P., Cinková, S., Fučíková, E., Mikulová, M., Pajas, P., Popelka, J., Semecký, J., Šindlerová, J., Štěpánek, J., Toman, J., Urešová, Z., and Žabokrtský, Z. (2011). Prague Czech-English Dependency Treebank 2.0.

[5] Hajičová, E. (2008). What We Are Talking about and What We Are Saying about It. In Gelbukh, A. F., editor, *Computational Linguistics and Intelligent Text Processing*. Springer Berlin /Heidelberg, Berlin, Heidelberg.

[6] Hajičová, E. and Sgall, P. (2003). Dependency Syntax in Functional Generative Description. *Dependenz und Valenz–Dependency and Valency*, 1:570–592.

[7] Kettnerová, V. (2012). *Lexikálně-sémantické konverze ve valenčním slovníku*. PhD thesis, Charles University, Prague, Czech Republic.

[8] Kettnerová, V., Lopatková, M., and Bejček, E. (2012). The Syntax-Semantics Interface of Czech Verbs in the Valency Lexicon. In Fjeld, R. and Torjusen, J., editors, *Proceedings of the 15th EURALEX International Congress*, pages 434–443, Oslo, Norway. Department of Linguistics and Scandinavian Studies, University of Oslo.

[9] Kingsbury, P. and Palmer, M. (2002). From Treebank to Propbank. In *Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC-2002)*, pages 1989–1993. Citeseer.

[10] Levin, B. (1993). *English Verb Classes and Alternations*. The University of Chicago Press, Chicago and London.

[11] Lopatková, M., Kettnerová, V., Bejček, E., Skwarska, K., and Žabokrtský, Z. (2012). VALLEX 2.6.

[12] Mikulová, M., Bémová, A., Hajič, J., Hajičová, E., Havelka, J., Kolářová, V., Kučová, L., Lopatková, M., Pajas, P., Panevová, J., Ševčíková, M., Sgall, P., Štěpánek, J., Urešová, Z., Veselá, K., and Žabokrtský, Z. (2007). Annotation on the Tectogrammatical Level in the Prague Dependency Treebank. Technical Report 3.1, ÚFAL, Charles University.

[13] Ruppenhofer, J., Ellsworth, M., Petruck, M. R., Johnson, C. R., and Scheffczyk, J. (2006). Framenet II: Extended theory and practice.

[14] Šindlerová, J. and Bojar, O. (2009). Towards English-Czech Parallel Valency Lexicon via Treebank Examples. In *Eighth International Workshop on Treebanks and Linguistic Theories*, pages 185–195.

[15] Urešová, Z. (2011a). *Valence sloves v Pražském závislostním korpusu.* Studies in Computational and Theoretical Linguistics. Ústav formální a aplikované lingvistiky, Praha, Czechia.

[16] Urešová, Z. (2011b). *Valenční slovník Pražského závislostního korpusu (PDT-Vallex).* Studies in Computational and Theoretical Linguistics. Ústav formální a aplikované lingvistiky, Praha, Czechia.

[17] Urešová, Z., Fučíková, E., Hajič, J., and Šindlerová, J. (2013). An Analysis of Annotation of Verb-Noun Idiomatic Combinations in a Parallel Dependency Corpus. To appear in the Proceedings from The 9th Workshop on Multiword Expressions, Workshop at NAACL 2013.