

Inherently Pronominal Verbs in Czech: Description and Conversion Based on Treebank Annotation

Zdeňka Urešová

Eduard Bejček

Jan Hajič

Charles University in Prague, Faculty of Mathematics and Physics

Institute of Formal and Applied Linguistics

Malostranské nám. 25

11800 Prague 1, Czech Republic

{uresova,bejcek,hajic}@ufal.mff.cuni.cz

Abstract

This paper describes results of a study related to the PARSEME Shared Task on automatic detection of verbal Multi-Word Expressions (MWEs) which focuses on their identification in running texts in many languages. The Shared Task’s organizers have provided basic annotation guidelines where four basic types of verbal MWEs are defined including some specific subtypes. Czech is among the twenty languages selected for the task. We will contribute to the Shared Task dataset, a multilingual open resource, by converting data from the Prague Dependency Treebank (PDT) to the Shared Task format. The question to answer is to which extent this can be done automatically. In this paper, we concentrate on one of the relevant MWE categories, namely on the quasi-universal category called “Inherently Pronominal Verbs” (IPronV) and describe its annotation in the Prague Dependency Treebank. After comparing it to the Shared Task guidelines, we can conclude that the PDT and the associated valency lexicon, PDT-Vallex, contain sufficient information for the conversion, even if some specific instances will have to be checked. As a side effect, we have identified certain errors in PDT annotation which can now be automatically corrected.

1 Introduction

Although Multi-Word Expressions (MWEs) attract the attention of more and more NLP researchers, as stated in Rosén et al. (2015), there is no consensus both on MWEs annotation and on what constitutes a MWE. This complicates the research of MWEs based on annotated corpora and language resources. To remedy this situation, the COST network PARSEME¹ (Savary et al., 2015) concentrates on the study of MWEs and their annotation in treebanks aiming at building a set of standardized annotation principles, corpora and evaluation metrics.

In the framework of PARSEME, a Shared Task on automatic detection of verbal Multi-Word Expressions was established in order to provide a multilingual open resource to be available to the NLP community. This initiative runs from 2015

to 2017. There are about twenty corpus contributors to the Shared Task. The task covers languages of different language families. Languages are divided into four language groups of comparable sizes: Germanic, Romance, Slavic and other. Common standardized annotation guidelines have been developed which try to define common principles of verbal MWE annotation, while also taking language specifics into account (Vincze et al., 2016). The guidelines summarize the properties of verbal MWEs and provide basic annotation rules for them. Various types of verbal MWEs as identified by previous research have been classified into seven groups: light verb constructions (LVC), idioms (ID), and then possibly verb particle combinations (VPC), inherently pronominal verbs (IPronV) and inherently prepositional verbs (IPrepV) if these three quasi-universal categories are applicable in the language, possibly other language specific category, and other verbal MWEs (OTH).

In our paper, we concentrate on the inherently pronominal verbs (IPronV) category. The paper is structured as follows: In Section 2, the Czech data (PDT) and the valency lexicon PDT-Vallex are presented. In Section 3, the category of inherently pronominal verbs (IPronV) is described focusing on Czech language specifics. In Section 4, we focus on the relation of the specification of the IPronV category for the Shared Task and the PDT-Vallex and PDT annotation, which then forms the starting point for the conversion procedure into the format of the PARSEME Shared Task. Section 5 concludes the paper.

2 Czech data and lexicon

For our study, we use data from the Czech annotated corpus, the PDT, as described in Sect. 2.1, and from the Czech valency lexicon PDT-Vallex (Sect. 2.2).

¹<http://typo.uni-konstanz.de/parseme>

2.1 The Prague Dependency Treebank

The Prague Dependency Treebank 2.0 (Hajič et al., 2006) published by the Linguistic Data Consortium² contains Czech written texts with complex and interlinked morphological, syntactic and complex semantic annotation.³ Its annotation scheme is based on the formal framework called Functional Generative Description (FGD) (Sgall et al., 1986), which is dependency-based with a “stratificational” (layered) approach to a systematic description of a language. The annotation contains interlinked surface dependency trees and deep syntactic/semantic (*tectogrammatical*) trees. Valency is one of the core FGD concepts, used on the deep layer (Panevová, 1974; Panevová, 1994). We shall note that each verb occurrence at the tectogrammatical level of annotation contains a manually assigned link (in a form of a unique frame ID) to the corresponding valency frame in the valency lexicon (Sect. 2.2).

The PDT has been extended in its versions PDT 2.5 (Bejček et al., 2012) and subsequently in PDT 3.0⁴ by adding, e.g., extensive MWE annotation. However, since we are focusing on IPronV in this paper, we have in fact not used this extension, which concerns other (mostly nominal) types of MWEs.

2.2 PDT-Vallex – Czech valency lexicon

The Czech valency lexicon, called PDT-Vallex is publicly available⁵ as a part of the PDT family of treebanks; for details, see Urešová (2011), Dušek et al. (2014) and Urešová et al. (2016), which we very briefly summarize here. As such, it has been designed in close connection with the specification of the treebank annotation. Each verb occurrence in the PDT is linked to a specific verb valency frame in the valency lexicon.

Each valency entry in the lexicon contains a headword, according to which the valency frames are grouped, indexed, and sorted. The valency frame consists of valency frame members (slots) and their labels, the obligatoriness feature for each member and the required surface form of valency frame members. Any specific lexical realization of a particular valency frame is exempli-

fied by an understandable fragment of a Czech sentence. Valency frame members are labeled by *functors* based on the FGD theory (ACT for Actor, or first argument, PAT for Patient or 2nd argument, ADDRessee, EFFECT and ORIGIN for the remaining core argument, and any other functor if deemed obligatory). Notes help to delimit the meaning (verb sense) of the individual valency frames within one valency lexicon entry. In the notes, synonyms, antonyms and aspectual counterparts are often found as additional hints to distinguish among the individual valency frame senses. An example of a valency lexicon entry for *tolerovat* (lit. *tolerate*) is in Fig. 1.

tolerovat	
tolerovat ¹ _{7X,6X} ACT(1) PAT(4,;ž;:c) ?ADDR(3)	(snášet, trpět) • toleroval jí její zálety; t. nedostatky
tolerovat ² _{1X} ACT(1) PAT(4)	(snášet) • Lidé byli ochotní ho kvůli krátkodobým ziskům tolerovat.

Figure 1: A simple PDT-Vallex entry with two senses (lit. *tolerate*): to tolerate¹ *sth*.PAT to *sb*.ADDR, to tolerate² *sb*.PAT

PDT-Vallex 2.0 which we have used in our work contains 11,933 valency frames for 7,121 verbs.

3 Inherently pronominal verbs

The PARSEME Shared Task general guidelines (Vincze et al., 2016) define the IPronV category as a specific quasi-universal⁶ verbal MWE category.

We use the guidelines for IPronV identification (Candito and Ramisch, 2016) where the basic rules are described. The guidelines divide verbs with a pronominal clitic into several groups. The first group of IPronV never occurs without the clitic – the clitic must co-occur with the verb, such as:

- cs: *bát se* (lit. *be afraid*)
- fr: *s'évanouir* (lit. *faint*)
- pl: *dowiedzieć się* (lit. *find out*)
- pt: *queixar-se* (lit. *complain*)

The second group of IPronV contains such verbs that might occur without the clitic, but with a different meaning:

²<http://www ldc.upenn.edu/LDC2006T01>

³<https://ufal.mff.cuni.cz/pdt2.0>

⁴<http://hdl.handle.net/11858/00-097C-0000-0023-1AAF-3>

⁵<http://hdl.handle.net/11858/00-097C-0000-0023-4338-F>

⁶*quasi-universal* = not found / defined for all languages, as opposed to *light verb constructions* (LVC) and *idioms* (ID), which are universal for all languages within the PARSEME Shared Task.

- cs: *hledět si* (lit. *mind sth*) vs. *hledět* (lit. *watch*)
- fr: *s'agir* (lit. *be*) vs. *agir* (lit. *act*)
- pl: *radzić sobie* (lit. *manage*) vs. *radzić* (lit. *advise*)
- pt: *encontrar-se* (lit. *be*) vs. *encontrar* (lit. *meet*)

The guidelines also list several other groups of pronominal verbs defined by an occurrence in a predominantly different syntactic behavior (reciprocals, reflexives, possessive reflexives, synthetic passives, impersonal and inchoative), which are *NOT* considered verbal MWEs (IPronV) unless their meaning has shifted substantially.

Given the complexity of this kind of verbal MWEs, the guidelines for the annotation of IPronV contain a detailed suit of tests for the proper annotation of IPronV. These tests are in the form of a binary decision tree that shows how to apply the tests in order to distinguish which pronominal verb occurrence has to be annotated as verbal MWEs and which should not. For example, test No. 8 distinguishes between a reciprocal use with plural subject and a real inherently pronominal construction:

*Is it possible to remove the reflexive particle and replace the coordinated subject (A and B) or plural subject (A.PL) by a singular subject (A or A.PL) and a singular object, often introduced by to/with (B or A.PL), without changing the pronominal verb's meaning? If yes, it is not IPronV.*⁷

3.1 Czech verbs with reflexive particles

The issue of Czech reflexives has been described by many scholars, e.g., Štícha (1981), Panevová (1999) or Panevová (2007), from diverse point of views. For example, in Kettnerová and Lopatková (2014) Czech reflexive verbs are dealt from the lexicographic point of view and a proposal for their lexicographic representation is formulated. Although reflexives are the topic of Czech theoretical (Panevová and Mikulová, 2007; Oliva, 2001) as well as computational linguistic papers (Petkevič, 2013; Oliva, 2003), as far as we know, there is no unified theoretical description of this language phenomenon. We believe the reason is the complexity of this ambiguous phenomenon since the Czech reflexive particle *se* or *si* can be used

⁷Candito and Ramisch (2016), page 7

both as formal morphological means for word-formation (e.g., reflexivization) and as syntactic means for specific syntactic structures (reflexivity, reciprocity, diatheses). Specifically, *se* is (a) a short (clitic) form of the pronoun *sebe* (lit. all of *itself, myself, yourself, herself, himself, ourselves, yourselves, themselves*) in accusative case, or (b) a reflexive particle for regular formation of passive constructions, particle for “frozen” constructions where it diachronically became part of the verb lexeme (except it is not written together with the verb form; it can be placed quite far from it in a sentence), as well as (c) the reflexivization particle for certain additional types of constructions, such as medio-passive construction of disposition *it reads well* which is expressed in Czech by adding this particle to the verb form (*čte se to dobře*).⁸

3.2 Inherently pronominal verbs in the PDT-Vallex and in the PDT

As has been already mentioned, we are investigating whether the information present in the PDT-Vallex (and in the PDT) can be used for determining the IPronV class. Although the detailed information about specific types of pronominal verbs is not explicitly captured in the PDT-Vallex, it does contain information related to the use of reflexive particles *se* or *si* in Czech. Moreover, the lexicon is linked to the PDT, so each corpus occurrence can be related to the lexicon (and vice versa).

The formal indicator that has been used in the PDT-Vallex to denote “reflexivization” (in the sense used in the PDT and PDT-Vallex annotation, see Mikulová et al. (2006)) is the addition of the particle *se* or *si* to the lemma (entry headword). Therefore, there might be up to three different headwords for each verb lemma in the PDT-Vallex: one without any such particle, one with *se* and one with *si*.⁹

Pronominal *se/si* is the only case of MWE captured in the PDT-Vallex as a headword, which illustrates its specificity in Czech. Czech does not display other similar phenomena such as phrasal verbs in English (*look up, run away* etc.).¹⁰

⁸Moreover, *se* is also a vocalized preposition used with the instrumental case, corresponding to English *with*.

⁹Just for completeness, there can never be both particles at the same time.

¹⁰However, LVCs and IDs do exist in Czech and they can also be identified in PDT-Vallex; see Fig. 2 and its description at the end of this section.

In addition and to our advantage here, PDT-Vallex stores different verb senses separately, as different valency frames under the same headword.¹¹

When we applied the specific tests for annotation of IPronV and went through the suggested decision tree step by step, we have determined that the first three questions (inherent reflexives, i.e., reflexives tantum), inherent reflexives due to different senses (i.e., derived reflexives), and inherent reflexives with a different subcategorization than the verb without the particle (i.e., derived reflexives) are easily answered by simply testing the existence of the *se* or *si* particle in the headword of a particular valency frame. In other words, *all* valency frames the headword of which contains the *se* or *si* particle will be marked as IPronV.

We have then analyzed the follow-up tests in the guidelines. These tests, similarly to the Plural/Coordination test shown earlier, test whether the occurrence of the verb construction is rather of a syntactic nature (deagentives etc.), and if YES, it disallows to annotate it as IPronV. However, it was found that since PDT-Vallex abstracts from—or generalizes over—such constructions, keeping only the basic (canonical, active voice) valency frame, we can in fact rely on the *se* or *si* indicators at the headword also for these special cases. In other words, diatheses are not explicitly present in the PDT-Vallex, they are assumed to be formed by regular derivation processes (such as reflexive or periphrastic passivization, reciprocalization, etc.) on the basis of the canonical valency frame as recorded in PDT-Vallex. Since the links from the PDT corpus to the individual valency frames in PDT-Vallex also abstract from such diathetical transformations, we do not have to apply such tests to the PDT-Vallex entries when distinguishing IPronV.

To summarize, we have determined that due to the way PDT-Vallex is structured and linked to the corpus, the only necessary indication that the phrase should be marked as IPronV is that the valency frame it is linked to has a headword with the *se* or *si* particle. In other words, albeit without knowing it, the annotators and creators of PDT-Vallex have already built in the IPronV MWE type in the lexicon using the *se/si* indicator.

¹¹The valency frames for different verb senses for each headword have often different syntactic and semantic description—such as the number of arguments, their surface realization etc.—but they might be identical.

Statistics for 1580 inherently pronominal verbs as found in the PDT-Vallex are given in Table 1.

Type of IPronV	Particle	
	<i>se</i>	<i>si</i>
reflexive tantum verbs	587	98
derived reflexives	743	152

Table 1: Statistics on 1580 inherently pronominal verbs in PDT-Vallex. Reflexive tantum verbs: clitic is compulsory; derived reflexives: absence of the clitic changes the sense.

Table 2 shows numbers for 72 verbs (headwords in PDT-Vallex) where we expect the annotation to contain overlapping labels IPronV and one of ID or LVC for at least one frame. The number of all ID/LVC frames belonging to pronominal verbs (headwords) is 172.

Type of overlap	No. of headwords
ID only	58
LVC only	9
LVC and ID	5

Table 2: Statistics on verbs potentially overlapping IPronV and ID/LVC annotation.

An example of valency lexicon entry for the verb (headword) *dělat si* with all its valency frames (senses) is displayed in Fig. 2. The first and last frame describes a MWE of inherently pronominal verb meaning, and each occurrence in the corpus can be thus labeled IPronV. All the other frames are examples of an embedded MWE, since on top of being an IPronV, they are also of the LVC category (those having one of the arguments labeled CPHR) or of the ID (idiom) category (those having one of the arguments labeled DPHR). In these seven cases two embedded MWE can be labeled at once: IPronV and either ID or LVC.

4 Conversion of Czech data

Based on the results of the investigation described in Sect. 3.2, we can conclude that the category of IPronV as defined in the guidelines for the PARSEME Shared Task corresponds to such verbs in the PDT whose tectogrammatical lemma contains *se* or *si* in a form of a “word with spaces”.

dělat si	
dělat si ¹ _{3x,2x} ACT(1) PAT(4;c)	(činit si) • dělal si výlety do okolí; dělá si, co chce
dělat si ² _{3x,4x} ACT(1) CPHR((naděje,nárok,představa...).4)	(činit si) • dělali si představu o velikosti souboru
dělat si ³ _{15x} ACT(1) CPHR((obava.starost...).4)	• Investoři si dělali starost.; Maminka si dělala obavy o syna.
dělat si ⁴ ACT(1) DPHR(hlava:S4[těžký.#]) PAT(z+2)	• dělat si z toho těžkou hlavu
dělat si ⁵ _{2x,2x} ACT(1) DPHR(nic.4) PAT(z+2)	• nic si z toho nedělejte
dělat si ⁶ _{2x,3x} ACT(1) DPHR(legrace.S4;blázen.P4) PAT(z+2)	• dělala si z nich legraci; dělal si z ní blázný
dělat si ⁷ _{2x} ACT(1) DPHR(iluze.P4)	• nedělejte si iluze
dělat si ⁸ _{2x} ACT(1) DPHR(hlava.S4;hlava.S4[velký.#]) PAT(s+7.z+2)	• Student si s tím hlavu nedělá.; Zaměstnanci si z toho velkou hlavu nedělají.
dělat si ⁹ _{1x} ACT(1) PAT(l.c) ADDR(s+7)	(zacházet) • Společnost si s poštou dělala, co chtěla.

Figure 2: An example of PDT-Vallex entry with several senses of the verb *dělat* (lit. *do/make*) in which the particle *si* has to co-occur.

Translation of Czech examples:

*dělat si*¹: *he's making short trips in the neighborhood;*

he does whatever he wants

*dělat si*² (*představu=imagine*): *they were imagining the size of the file*

*dělat si*³ (*starost=worry*): *the investors were worried;*

mother was worried about son

*dělat si*⁴ (*těžkou hlavu=heavy head*): *he did worry*

*dělat si*⁵ (*ne- nic z=not nothing out-of*): *don't worry about it*

*dělat si*⁶ (*legraci, blázný=fun, fools*): *she was making fun of them; he was making monkey out of her*

*dělat si*⁷ (*ne- iluze=not illusions*): *let's don't delude ourselves*

*dělat si*⁸ (*velkou hlavu=big head*): *student is not worried about it; employees do not worry much about it*

*dělat si*⁹ (*s=with*): *the company did whatever they liked with the mail*

However, having the tectogrammatical annotation of the PDT linked to the surface dependencies, we have also checked the lexicon annotation against the corpus not only through the reference linking the PDT's tectogrammatical annotation to PDT-Vallex, but also against the surface dependency annotation.

We worked with a hypothesis that all the IPronV should be linked to a surface verb and a separate node for the particle (*se* or *si*), and that the syntactic function of the *se* or *si* node should be labeled as AuxT. Analytical function AuxT is assigned to the particles *se* or *si* in case the verb sense without them does not exist, which to a large extent also corresponds to the IPronV property at

the surface syntactic level (Hajič et al., 2004).¹² We found that in 93.1% of the occurrences, this is indeed the case, but there are more than 700 cases where the syntactic relation was different (not AuxT). After investigating a sample of those, we found that they were errors (such as holding the Adv, Obj, AuxO or AuxR label) in the surface dependency annotation. These cases will not be used for the conversion to the PARSEME Shared Task dataset, unless further investigation can prove that they are indeed all just surface annotation errors in the original data.

5 Conclusions

We have compared the annotation of verbal entries in the PDT (and PDT-Vallex) with the PARSEME Shared Task guidelines for inherently pronominal verbs. The main conclusion is that albeit annotated independently, the PDT/PDT-Vallex annotation covers all IPronV categories relevant for Czech as defined in the guidelines.

By a relatively simple conversion process we have also checked the annotation at the surface syntactic dependency annotation level of the PDT and found a few mismatches. At this time, these mismatches seem to be mostly errors of the surface dependency level annotation in the PDT.¹³

Acknowledgements

The work described herein has been supported by the grant GP13-03351P of the Grant Agency of the Czech Republic, by the grant LD14117 of the Ministry of Education, Youth and Sports of the Czech Republic (MEYSCR) and by the LINDAT/CLARIN Research Infrastructure projects, LM2010013 and LM2015071 funded by the MEYSCR. The project also accepted support from the COST Action No. IC1207 "PARSEME", funded by the European Commission. It has also been using language resources developed and distributed by the LINDAT/CLARIN project LM2015071 (<http://lindat.cz>).

¹²Except it might be elided, shared among several verb tokens in coordination etc., so that it is not that "reliable" indicator as the presence of the reflexive particle in the PDT-Vallex headword; for details, see Hajič (1998).

¹³The conversion process should be finished (including all other types of verbal MWEs as defined by the PARSEME Shared Task guidelines, which will be described in a followup publication) by the time of the workshop and we will provide full statistics.

References

- Eduard Bejček, Jarmila Panevová, Jan Popelka, Pavel Straňák, Magda Ševčíková, Jan Štěpánek, and Zdeněk Žabokrtský. 2012. Prague dependency treebank 2.5 – a revisited version of PDT 2.0. In Martin Kay and Christian Boitet, editors, *Proceedings of the 24th International Conference on Computational Linguistics (Coling 2012)*, pages 231–246. Mumbai, India. IIT Bombay, Coling 2012 Organizing Committee.
- Marie Candito and Carlos Ramisch. 2016. Annotation guidelines for inherently pronominal verbs (PARSEME shared task on automatic detection of verbal MultiWord Expressions), <http://typo.uni-konstanz.de/parseme/images/shared-task/guidelines/PARSEME-ST-annotation-guidelines-iprovnv.pdf>.
- Ondřej Dušek, Jan Hajič, and Zdeňka Urešová. 2014. Verbal valency frame detection and selection in Czech and English. In *The 2nd Workshop on EVENTS: Definition, Detection, Coreference, and Representation*, pages 6–11, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Jan Hajič, Jarmila Panevová, Eva Buráňová, Zdeňka Urešová, Alevtina Bémová, Jan Štěpánek, Petr Pajas, and Jiří Kárník. 2004. Anotace na analytické rovině. Návod pro anotátory. Technical Report TR-2004-23, ÚFAL/CKL MFF UK, Prague.
- Jan Hajič. 1998. *Building a Syntactically Annotated Corpus: The Prague Dependency Treebank*. Karolinum, Charles University Press, Prague.
- Jan Hajič, Jarmila Panevová, Eva Hajičová, Petr Sgall, Petr Pajas, Jan Štěpánek, Jiří Havelka, Marie Mikulová, Zdeněk Žabokrtský, Magda Ševčíková Razímová, and Zdeňka Urešová. 2006. *Prague Dependency Treebank 2.0*. Number LDC2006T01. LDC, Philadelphia, PA, USA.
- Václava Kettnerová and Markéta Lopatková. 2014. Reflexive verbs in a valency lexicon: The case of czech reflexive morphemes. In Andrea Abel, Chiara Vettori, and Nascia Ralli, editors, *Proceedings of the XVI EURALEX International Congress: The User in Focus*, pages 1007–1023, Bolzano/Bozen, Italy. EURAC research, EURAC research.
- Marie Mikulová, Alevtina Bémová, Jan Hajič, Eva Hajičová, Jiří Havelka, Veronika Kolářová, Lucie Kučová, Markéta Lopatková, Petr Pajas, Jarmila Panevová, Magda Razímová, Petr Sgall, Jan Štěpánek, Zdeňka Urešová, Kateřina Veselá, and Zdeněk Žabokrtský. 2006. Annotation on the tectogrammatical level in the Prague Dependency Treebank. Annotation manual. Technical Report 30, Prague, Czech Rep.
- Karel Oliva. 2001. Reflexe reflexivity reflexiv. *Slovo a slovesnost*, 62(3):200–207.
- Karel Oliva. 2003. Linguistics-based PoS-tagging of Czech: disambiguation of se as a test. In *Contributions of the 4th European Conference on Formal Description of Slavic Languages*, pages 299–314. Postdam University, Germany.
- Jarmila Panevová and Marie Mikulová. 2007. On reciprocity. *The Prague Bulletin of Mathematical Linguistics*, (87):27–40.
- Jarmila Panevová. 1974. On verbal Frames in Functional Generative Description. *Prague Bulletin of Mathematical Linguistics*, 22:3–40.
- Jarmila Panevová. 1994. Valency frames and the meaning of the sentence. *The Prague School of Structural and Functional Linguistics*, 41:223–243.
- Jarmila Panevová. 1999. Česká reciproční zájmena a slovesná valence. *Slovo a slovesnost*, 60:91–100.
- Jarmila Panevová. 2007. Znovu o reciprocitě. *Slovo a slovesnost*, 68:269–275.
- Vladimír Petkevič. 2013. Formal (morpho)syntactic properties of reflexive particles se, si as free morphemes in contemporary Czech. In *Proceedings of the 7th International Conference SLOVKO 2013*, pages 206–216. Slovenska akademia vied, Slovakia.
- Victoria Rosén, Gyri Losnegaard, Koenraad De Smedt, Eduard Bejček, Agata Savary, Adam Przepiórkowski, Petya Osenova, and Verginica Mititelu. 2015. A survey of multiword expressions in treebanks. In *14th International Workshop on Treebanks and Linguistic Theories (TLT 2015)*, pages 179–193, Warszawa, Poland. IPIAN, IPIAN.
- Agata Savary, Manfred Sailer, Yannick Parmentier, Michael Rosner, Victoria Rosén, Adam Przepiórkowski, Cvetana Krstev, Veronika Vincze, Beata Wójtowicz, Gyri Smørdal Losnegaard, Carla Parra Escartín, Jakub Waszczuk, Matthieu Constant, Petya Osenova, and Federico Sangati. 2015. PARSEME – PARSing and Multiword Expressions within a European multilingual network. In *7th Language & Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics (LTC 2015)*, Poznań, Poland, November.
- Petr Sgall, Eva Hajičová, and Jarmila Panevová. 1986. *The meaning of the sentence in its semantic and pragmatic aspects*. D. Reidel, Dordrecht.
- Zdeňka Urešová, Eva Fučíková, and Jana Šindlerová. 2016. Czengvallex: a bilingual czech-english valency lexicon. *The Prague Bulletin of Mathematical Linguistics*, 105:17–50.
- Zdeňka Urešová. 2011. *Valenční slovník Pražského závislostního korpusu (PDT-Vallex)*. Studies in Computational and Theoretical Linguistics. Ústav formální a aplikované lingvistiky, Praha, Czechia.
- Veronika Vincze, Agata Savary, Marie Candito, and Carlos Ramisch. 2016. Annotation guidelines for the PARSEME shared task on automatic detection of verbal Multi-Word Expressions, version 5.0, <http://typo.uni-konstanz.de/parseme/images/shared-task/guidelines/PARSEME-ST-annotation-guidelines-v5.pdf>.
- František Štícha. 1981. K syntakticko-sémantické konkurenci aktivních, participiálních a reflexivních konstrukcí. *Slovo a slovesnost*, 42:183–192.