

# Tools for Building an Interlinked Synonym Lexicon Network

Zdeňka Urešová, Eva Fučíková, Eva Hajičová, Jan Hajič

Charles University, Faculty of Mathematics and Physics

Institute of Formal and Applied Linguistics

Malostranské nám. 25

11800 Prague 1, Czech Republic

{uresova,fucikova,hajicova,hajic}@ufal.mff.cuni.cz

## Abstract

This paper presents the structure, features and design of a new interlinked verbal synonym lexical resource called CzEngClass and the editor tool being developed to assist the work. This lexicon captures cross-lingual (Czech and English) synonyms, using valency behavior of synonymous verbs in relation to semantic roles as one of the criteria for defining such interlingual synonymy. The tool, called Synonym Class Editor - SynEd, is a user-friendly tool specifically customized to build and edit individual entries in the lexicon. It helps to keep the cross-lingual synonym classes consistent and linked to internal as well as to well-known external lexical resources. The structure of SynEd also allows to keep and edit the appropriate syntactic and semantic information for each Synonym Class member. The editor makes it possible to display examples of class members' usage in translational context in a parallel corpus. SynEd is platform independent and may be used for multiple languages. SynEd, CzEngClass and services based on them will be openly available.

**Keywords:** Lexical Resource, Parallel Corpus, Semantics, Syntax, Synonymy, Valency

## 1. Introduction

We present a demonstration of our lexicon editor, called **SynEd**, for creating an interlinked multilingual (for the time being, bilingual) lexical resource—a contextually-based synonym lexicon of verbs based on their syntactic and semantic behavior in (bilingual) context. We also present the design and structure (scheme) of the lexicon dataset proper.

The lexicon, under the working name “CzEngClass” (Urešová et al., 2018), is being built “bottom-up”, using a parallel bilingual corpus with a rich manual annotation (the Prague Czech-English Dependency Treebank, version 2.0), its associated valency lexicons, i.e., EngVallex, PDT-Vallex and CzEngVallex, and it is being linked to other external lexical resources—VALLEX, FrameNet, VerbNet, PropBank and Czech and English WordNets (Sect. 3.).

One of the crucial parts of the “CzEngClass” project was to establish a scheme for this resource (Sect. 4.), taking into account not only its intended multilinguality, but also links to the initial bilingual parallel corpus (and more corpora in the future) and the existing relevant lexical resources. The semantic role labels used for defining the individual synonym classes are based mostly on FrameNet (and in some cases on VerbNet), while the argument labels and their morphosyntactic behavior are taken from the valency lexicons used for the PCEDT annotation. One of the main principles for the design of “CzEngClass” follows A. Kilgariff’s idea of corpus - dictionary linkage (cf. the PDIC and PCID model described in (Kilgariff, 2005)), so we are strictly keeping references to all of the used lexical resources (internal and external) as well as to the corpus examples influencing the class divisions.

Since the goal of the project is to identify contextually-based Czech and English synonyms, each verb has to be first broken down to senses. The initial set of senses for both Czech and English has been taken from the Czech and English valency lexicons, since they have been determined

during the creation of these lexicons and linked manually to each occurrence of the verb in the parallel corpus.

**SynEd**, the editor for “CzEngClass” (Sect. 4.), is currently used as a standalone application, with links to all the referenced external resources that can be immediately accessed directly in **SynEd**, through third-party applications and/or through web browsers. Currently, there are 60 synonym classes with approx. 470 English and over 1000 Czech verbs (verb senses) included in CzEngClass.<sup>1</sup> A selection and/or creation of a web-based customized interface for browsing and searching will follow and is part of future work (Sec. 5.). The editor as well as the associated data will be publicly available under a CC license.

## 2. Related Work

When designing the “CzEngClass” lexicon as well as the editor, we were looking for an existing annotation tool for a similar type of lexicon(s) we could possibly adapt. We concentrated on those that allow working with corpora, since that is also the way we approach building the lexicon.

Lexicons have been built using software tools (and corpora) since the 1980s, mainly at publishers, such as (Ahlsweide, 1985); such efforts are summarized in (Teubert, 2007). We have considered many other existing tools, either standalone or available as web services and applications. *Lexicon Creator* is a tool designed to help developers produce lexical data for its use in a variety of linguistic applications. According to (Fontenelle et al., 2008), *Lexicon Creator* enables to work on existing wordlists derived either directly from corpora or from previously created wordlist data. *Lexicon Builder*, available as web service (Parai et al., 2010), aims at automated methods to compile custom lexicons from BioPortal ontologies. *CoBaLT* (Kenter et al., 2012), is a web-based editor optimized for work with large

<sup>1</sup>The final size should cover at least the current Czech and English valency lexicons, i.e. about 15,000 verbs (verb senses); it is very hard to estimate the number of classes created in the end.

datasets and to produce historical lexica. *Dicet* (Gader et al., 2012) is a knowledge-based, tailor-made lexical graph editor and browser that allows lexicographers to browse through the lexical network and directly expand and revise it. *DECFC*,<sup>2</sup> a specialized dictionary editor (Decary and Lapalme, 1990), provides a multi-windowing environment that enables the simultaneous execution of different processes on different parts of the screen. A database schema for developing and maintaining Japanese linguistic resources (Asahara et al., 2002) is a stand-off framework combining XML and a relational database. SIL's<sup>3</sup> latest version (8.3) of *FLEX* (FieldWorks Language Explorer) is a next specific program designed to assist linguists in collecting, managing and publishing linguistic data.<sup>4</sup> *FLEX* features powerful bulk editing tools and a large number of built-in fields.

For the resources we are linking CzEngClass to, there are also several tools. A specific editor - *Cornerstone* (Choi et al., 2010a) - has been specifically customized to create and edit frameset files for PropBank project.<sup>5</sup> One of the biggest advantages of *Cornerstone* is that it accommodates several languages (it was used for e.g., Arabic, Chinese, English, Hindi, and Korean). A semi-automatic *VerbaLex-FrameNet* linking tool (Materna, 2009; Materna and Pala, 2010; Materna, 2011; Materna, 2014) has been developed. This tool aims to build a core of Czech FrameNet.

All the above mentioned editors and tools are very sophisticated and useful, but rather specialized for the particular lexical resource. Those few exceptions, such as the Japanese lexical resource builder, are on the other hand too general and would need a substantial amount of customization, since for the CzEngClass lexicon we need to express more specific requirements. We have thus decided to write a new editor, reusing some parts that have been developed in the past for editing the valency lexicons and linking them to the associated treebanks. This new editor is called **SynEd**, and we describe it in (Sect. 4.).

### 3. Resources Used

Our approach to the development of a synonym lexicon for both NLP and linguistic studies builds on corpus examples with natural contexts. Therefore, we build our research on electronically accessible and richly annotated data, namely, on the Prague Czech-English parallel treebank and on the Prague Dependency Treebank (Hajič et al., 2006) valency lexicons (PDT-Vallex, EngVallex and CzEngVallex), as well as on other well-established lexical databases (e.g., FrameNet, VerbNet, Semlink, PropBank and Czech and English WordNets).

The core corpus resource is the Prague Czech-English Dependency Treebank (Hajič et al., 2012), which stores parallel PDT-style annotations (manual annotation of morphology, syntax and semantics) of English texts (Wall Street Journal part of Penn Treebank (Marcus et al., 1993)) and

<sup>2</sup>DECFC (Explanatory and Combinatory Dictionary of Contemporary French).

<sup>3</sup>Originally known as the Summer Institute of Linguistics, Inc.

<sup>4</sup><https://software.sil.org/fieldworks>

<sup>5</sup>For proper PropBank annotation another editor called *Jubilee* (Choi et al., 2010b) has been used.

their professional translation into Czech<sup>6</sup>. The PCEDT annotations capture linkage of the surface and deep syntactic layers; moreover, the deep layer contains verbal word sense labeling by keeping links of each verb occurrence to the appropriate valency frame in the associated valency lexicons, PDT-Vallex and EngVallex.

The main lexicon resources we build on are thus the following valency lexicons: EngVallex (Cínková et al., 2014),(Cínková, 2006), PDT-Vallex (Urešová et al., 2014), (Urešová, 2011) and also CzEngVallex (Urešová et al., 2015), (Urešová et al., 2016). These lexicons are based on the Functional Generative Description Valency Theory (FGDVT)<sup>7</sup>. In the lexicons, each entry has a headword with one or more valency frame(s). Every valency frame contains labeled arguments, their obligatoriness and the required surface form of valency frame members (arguments). PDT-Vallex contains about 12,000 valency frames for about 7,000 verbs; EngVallex, contains about 7,000 valency frames for almost 4,500 verbs. CzEngVallex links them across Czech and English using the automatic PCEDT corpus alignments (after manual pruning of erroneous alignments has been applied).<sup>8</sup>

Since CzEngClass aims at synonymy based on semantics, we also use the following lexical resources: FrameNet (Fillmore et al., 2003; Fillmore et al., 2003), VerbNet (Schuler, 2006), Semlink (Palmer, 2009; Bonial et al., 2012), PropBank (Palmer et al., 2005), Czech WordNet (Pala et al., 2011), (Pala and Smrž, 2004) and English WordNet (Miller, 1995; Fellbaum, 1998).<sup>9</sup> These resources are mainly being referred to by the newly built CzEngClass entries; in addition, FrameNet and VerbNet semantic roles are being consulted when defining any particular synonym class.

### 4. Lexicon Design

The lexicon groups translational verbal equivalents, i.e., verb senses, together both in Czech and English, originally represented as valency frames in the Czech and English valency lexicons, into "Synonym Classes." We call the synonymous senses in one class "Class Members."

Each class is assigned a common set of "Semantic Roles" (SRs) and the valency frame elements of each Class Member are mapped to SRs assigned to this set. This mapping is the crucial defining feature of the CzEngClass lexicon, since it defines, through the original argument mapping to their morphosyntactic features, the use of the arguments in text, and therefore the context in which the particular verb can be considered a synonym to the other, similarly defined (or "restricted") verbs (their senses).

Class Members are then further individually linked to the original lexicon (PDT-Vallex or EngVallex) and also to entry(ies) in all the additional resources used (if such rele-

<sup>6</sup>For further information see the project web page: <http://ufal.mff.cuni.cz/pcedt2.0>.

<sup>7</sup>For details on the FGDVT theory see e.g., (Panevová, 1977; Lopatková, 2010)

<sup>8</sup>The CzEngClass lexicon refers also to VALLEX (Lopatková et al., 2016), a much more elaborated lexicon, built on the same theoretical framework as PDT-Vallex. However, VALLEX it is not based on the PDT data.

<sup>9</sup><https://wordnet.princeton.edu/>

vant entries exists in them). More than one link to any of the additional resources can be included; in such a case, it means that the same or similar meaning has been found in the external lexicons, with an unclear distribution of senses. In other words, mappings between CzEngClass entries and the entries in the other lexical resources are not necessarily 1:1.

Table 1 shows a (simplified)<sup>10</sup> example. Both Czech and English verbs, determined to be synonymous in the particular sense defined by the PDT-Vallex and EngVallex links, are listed as this synonym class members. The semantic roles chosen for this class have been simplified from the FrameNet. Once selected, the arguments of the particular verb sense, as found in the valency lexicons, *must* be mapped to these semantic roles; however, if some of the SRs are not listed as an argument, the corresponding adjunct might be used, too. In the examples in Table 1, another phenomenon is displayed: in some cases, a semantic role can be expressed as one argument (e.g., PAT for *hear*, as in ... *they heard about it.PAT in local news*), or as two (part of the Phenomenon is expressed as PAT, part as EFF for *know*, as in - ... *he didn't know anything.EFF about him.PAT*). In such a case, all possibilities must be listed in the particular mapping field.

For such a lexicon, a stand-off XML schema has been developed and is used as the storage format. The XML document contains a header part, which gives local or remote reference to the external resources, lists all possible SRs (semantic roles) that are used in the lexicon (for consistency checking), etc. The body of the document contains the classes. Each class first lists the assigned SRs, and then the class members by verb lemmas, references to valency frames in PDT-Vallex and EngVallex defining the sense IDs, and references to external resources (FrameNet, VerbNet, PropBank, WordNets) for each class member. In addition, bookkeeping information is stored as well, such as annotator's ID, timestamps, etc. A (simplified) extract of the XML-formatted lexicon follows, for the class *build-budovat*:

```
<CzEngClass owner="EF">
... (header with SRs, arg. labels, lexicon URLs etc.)
<body>
<veclass lemma="budovat (v-w228f1)" id="vec00001">
  <commonroles>
    <role idref="vecroleAgent" spec=""/>
    <role idref="vecroleComponents" spec=""/>
    <role idref="vecroleCreated_Entity" spec=""/>
    <role idref="vecroleAssets_currency" spec=""/>
  </commonroles>
  <classnote/>
  <classmembers>
    <classmember status="not_touched" lang="cs"
      lexidref="pdtvallex"
      idref="PDT-Vallex-ID-v-w228f1"
      lemma="budovat" id="vec00001cm00001">
      <restrict/>
      <maparg>
        <argpair>
```

```
        <argfrom idref="vecargpdtACT"/>
        <argto idref="vecroleAgent"/>
      </argpair>
      <argpair>
        <argfrom idref="vecargpdtORIG"/>
        <argto idref="vecroleComponents"/>
      </argpair>
```

... (other arg pair mappings to SRs)

```
</maparg>
<cmnote/>
<extlex idref="pdtvallex">
  <links>
    <link idref="v-w228f1" lemma="budovat"/>
  </links>
</extlex>
<extlex idref="czengvallex">
  <links>
    <link idref="vw287f1p1" enid="ev-w384f1"
      enlemma="build" csid="v-w228f1" .../>
    <link idref="vw713f3p1" enid="ev-w918f3"
      enlemma="develop" csid="v-w228f1" .../>
```

... (more links to aligned valency frames)

```
    <link idref="vw2305f1p1" enid="ev-w2903f1"
      enlemma="set_up" csid="v-w228f1" .../>
```

```
  </links>
</extlex>
<examples> </examples>
<local_history>
  <local_event time_stamp="10.7.2017 22:38:45"
    type_of_event="edit ..." author="ZU"/>
</local_history>
</classmember>
```

... (more classmembers/synonyms)

```
</classmembers>
</veclass>
```

... (more classes of synonyms)

```
</body> </CzEngClass>
```

The **SynEd** editor (Fig. 1) that allows to view and edit the CzEngClass lexicon as described in the previous section, including management of external links to other lexicons and corpora (examples), has been created by a substantial rewrite of the original PDT-Vallex editor used for the annotation of the PDT (Hajič et al., 2006; Hajič et al., 2018). In order to allow fully offline but parallel work, it is a standalone perl program that runs on multiple OS platforms, supported by data synchronization over svn (and/or git). The editor allows editing of the XML document described above, and it understands all the external files (lexicons and corpora, as described in Sect. 3.). The annotator can thus see the original alignments across languages (from CzEng-Vallex and the PCEDT corpus), as well as the FrameNet and VerbNet classes taken as a basis for a particular synonym set. It also shows, for each synonym class member, only the relevant examples from the PCEDT corpus (that correspond to verb(s) in the other language, again based on the PCEDT Czech-English alignments). Among these example sentences, even though there are usually not that

<sup>10</sup>For space reasons, specific morphosyntactic restrictions and additional external links to PropBank and WordNets are not shown; the list of class members is also substantially shortened.

Class member	Roles			External Links		
	Cognizer	Phenomenon	Speaker	PDT/EngVallex	FrameNet	VerbNet
dozvědět se	ACT	PAT	ORIG	v-f1		
hear	ACT	PAT	ORIG	ev-f1	Hear	discover-84-1-1, see-30.1-1-1-1
know	ACT	PAT+EFF	ORIG	ev-f1	Awareness	consider-29.9-1-1
learn	ACT	PAT	ORIG	ev-f1	Becoming_aware	learn-14-2-1
doslýchat se	ACT	PAT+EFF	ORIG	ev-f1		
poučit se	ACT	PAT+EFF	ORIG	v-f1		
...	...	...	...	...	...	...

Table 1: Mappings and External Links for class “dozvědět se, hear, ..., poučit se, ...”

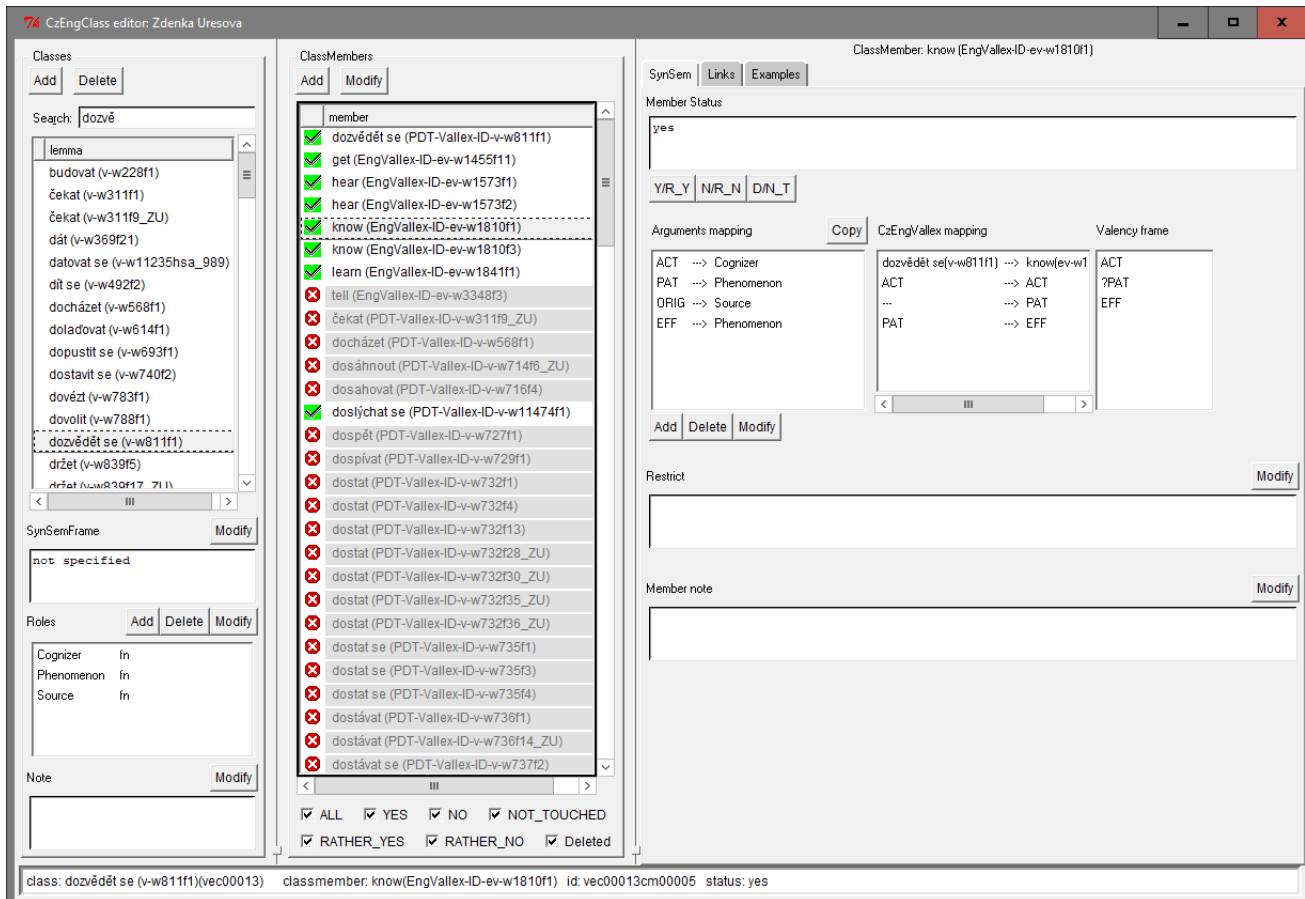


Figure 1: A screenshot of the SynEd editor

many in the relatively small PCEDT corpus, the annotator can mark the “best” (most illustrative) example sentence(s) from the corpus to be shown first to the future user once the lexicon is finished and made public.

As seen in Fig. 1, the editor’s screen contains three sections: list of all lemmas (allowing search for their class) and basic information for the selected class on the left, then a list of all class members (with PDT-Vallex/EngVallex IDs) in the middle, and then on the right information pertaining to the selected class member: argument mapping to SRs, status, additional restrictions and notes. The right-hand pane can be also switched to see and edit the Links to external resources (Fig. 2), or to see lists of Examples (using the tabs above the pane).

All external links are “clickable” in order to both simplify and speed up annotator’s work. The lexical re-

source links (tab “Links”, see also Fig. 2) are expanded to a full external or local URL and opened in a new browser tab. For example, the ID link for “draft,” which is part of the `create-26.4-1` VerbNet class, is expanded to `http://verbs.colorado.edu/verb-index/vn3.3/vn/create-26.4.php#create-26.4-1` and shown in the user’s new browser tab. Similarly, in the “Examples” tab, user can click on a “Show” button and the examples from PCEDT 2.0 are shown in the `tred`<sup>11</sup> viewer/editor, which must be installed locally.<sup>12</sup>

<sup>11</sup><https://ufal.mff.cuni.cz/tred>

<sup>12</sup>In the future, links to examples will possibly be converted to PML-TQ queries directed to `https://lindat.mff.cuni.cz/services/pmltq/#!/trebank/pcedt20`.

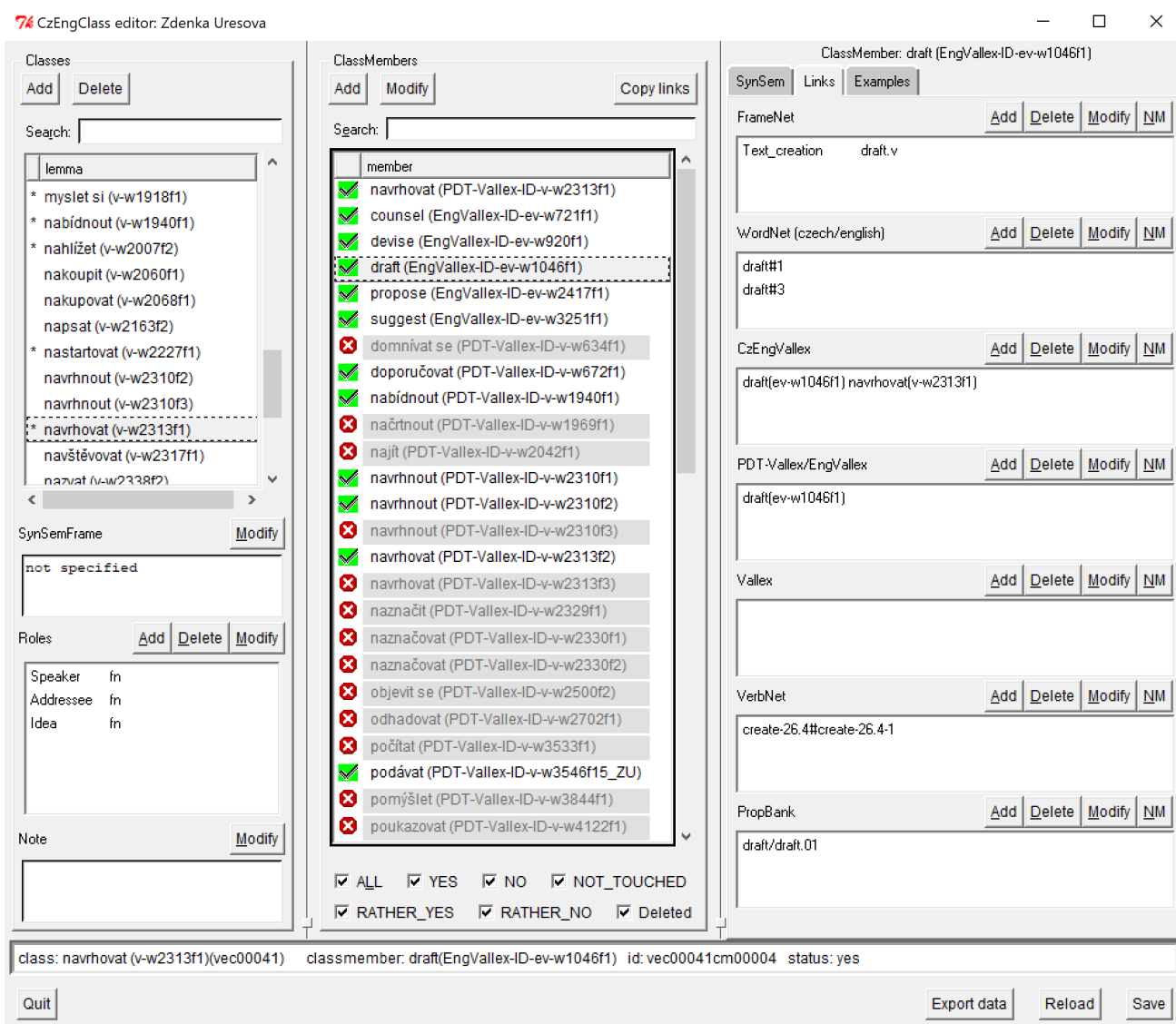


Figure 2: A screenshot of the Links to external resources tab in the **SynEd** editor

## 5. Future Work

This paper presented the design of the structure of the CzEngClass lexicon and the editor **SynEd** used for its editing and maintenance. The editor is currently being extensively tested in order to be used (simultaneously) by multiple annotators. Two annotators had worked with **SynEd** on the CzEngClass lexicon for almost a year so the editor has already been sufficiently tested and gradually improved, including time-saving as well as ergonomic features and is now ready to begin routine work.

We plan to investigate the relation of valency and semantic roles in more detail, also from their morphosyntactic realization point of view, and adjust our editor to such newly emerging requirements. We will continue to develop **SynEd** by improving its functionality and practicality through more feedback from future annotators.

In the future, we plan to explore cooperation with the OntoLex Community Group (Piasecki et al., 2017) in order to integrate our work into methods for representing linked lexical data.

We also plan to join the Cross-lingual FrameNet Group. We believe that our editor will be—possibly after some refinements—suitable for editing resources linked to FrameNet (and other lexical resources), including its deployment for other languages.

## Acknowledgments

This work has been supported by the grant No. GA17-07313S of the Grant Agency of the Czech Republic, and it uses resources hosted by the LINDAT/CLARIN Research Infrastructure, projects No. LM2015071 and CZ.02.1.01/0.0/0.0/16\_013/0001781, supported by the Ministry of Education of the Czech Republic. The last co-author has also been supported in part by the project VIA-DAT, No. DG16P02R019, of the Ministry of Culture of the Czech Republic.

## 6. Bibliographical References

- Ahlsvede, T. E. (1985). A Tool Kit for Lexicon Building. In *23rd Annual Meeting of the Association for Computational Linguistics*.
- Asahara, M., Yoneda, R., Yamashita, A., Den, Y., and Matsumoto, Y. (2002). Use of XML and Relational Databases for Consistent Development and Maintenance of Lexicons and Annotated Corpora. In *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC'02)*. European Language Resources Association (ELRA).
- Bonial, C., Feely, W., Hwang, J. D., and Palmer, M. (2012). Empirically Validating VerbNet Using SemLink. In *Seventh Joint ACL-ISO Workshop on Interoperable Semantic Annotation*, Istanbul, Turkey, May.
- Choi, J. D., Bonial, C., and Palmer, M. (2010a). Propbank Frameset Annotation Guidelines Using a Dedicated Editor, Cornerstone. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*. European Languages Resources Association (ELRA).
- Choi, J. D., Bonial, C., and Palmer, M. (2010b). Propbank Instance Annotation Guidelines Using a Dedicated Editor, Jubilee. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*. European Languages Resources Association (ELRA).
- Cinková, S. (2006). From PropBank to EngValLex: adapting the PropBank-Lexicon to the valency theory of the functional generative description. In *Proceedings of LREC 2006, Genova, Italy*.
- Decary, M. and Lapalme, G. (1990). An Editor for the Explanatory and Combinatory Dictionary of Contemporary French (DECFC). *Computational Linguistics*, 16(5):145–154.
- Fellbaum, C. (1998). *WordNet: An Electronic Lexical Database*. Language, Speech, and Communication. MIT Press, Cambridge, MA.
- Fillmore, C. J., Johnson, C. R., and L.Petruck, M. R. (2003). Background to FrameNet: FrameNet and Frame Semantics. *International Journal of Lexicography*, 16(3):235–250.
- Fontenelle, T., Cipollone, N., Daniels, M., and Johnson, I. (2008). Lexicon Creator: A Tool for Building Lexicons for Proofing Tools and Search Technologies. In Janet DeCesaris Elisenda Bernal, editor, *Proceedings of the 13th EURALEX International Congress*, pages 359–369, Barcelona, Spain, jul. Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.
- Gader, N., Lux-Pogodalla, V., and Polguère, A. (2012). Hand-Crafting a Lexical Network With a Knowledge-Based Graph Editor.
- Kenter, T., Erjavec, T., Žorga Dulmin, M., and Fiser, D. (2012). Lexicon Construction and Corpus Annotation of Historical Language with the CoBaLT Editor.
- Kilgariff, A. (2005). Putting the corpus into the dictionary. In *Proceedings of the second Meaning workshop*, pages 1–6, Trento, Italy, February.
- Lopatková, M., Kettnerová, V., Bejček, E., Vernerová, A., and Žabokrtský, Z. (2016). *Valenční slovník českých sloves VALLEX*. Nakladatelství Karolinum, Praha, Czechia.
- Lopatková, M. (2010). *Valency Lexicon of Czech Verbs: Towards Formal Description of Valency and Its Modeling in an Electronic Language Resource*. Habilitation thesis. Charles University in Prague, Praha, Czechia.
- Marcus, M. P., Santorini, B., and Marcinkiewicz, M. A. (1993). Building a Large Annotated Corpus of English: The Penn Treebank. *COMPUTATIONAL LINGUISTICS*, 19(2):313–330.
- Materna, J. and Pala, K. (2010). Using Ontologies for Semi-automatic Linking VerbaLex with FrameNet. In Nicoletta Calzolari (Conference Chair), et al., editors, *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta, may. European Language Resources Association (ELRA).
- Materna, J. (2009). Linking Czech Verb Valency Lexicon VerbaLex with FrameNet. In *Proceedings of 4th Language and Technology Conference*, pages 215–219, Poznań. Wydawnictwo Poznańskie Sp. z o.o.
- Materna, J. (2011). *Building FrameNet in Czech [online]*. Rigorózní práce, Masarykova univerzita, Fakulta informatiky, Brno.
- Materna, J. (2014). *Probabilistic Semantic Frames [online]*. Disertační práce, Masarykova univerzita, Fakulta informatiky, Brno.
- Miller, G. A. (1995). WordNet: A Lexical Database for English. *Commun. ACM*, 38(11):39–41, November.
- Pala, K. and Smrž, P. (2004). Building Czech Wordnet. *Romanian Journal of Information Science and Technology*, 7(1-2):79–88.
- Palmer, M., Gildea, D., and Kingsbury, P. (2005). The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics*, 31(1):71–106, March.
- Palmer, M. (2009). Semlink: Linking PropBank, VerbNet and FrameNet. In *Proceedings of the Generative Lexicon Conference*, page 9–15.
- Panevová, J. (1977). Verbal Frames Revisited. *The Prague Bulletin of Mathematical Linguistics*, (28):55–72.
- Parai, G. K., Jonquet, C., Xu, R., Musen, M. A., and Shah, N. H. (2010). The Lexicon Builder Web service: Building Custom Lexicons from two hundred Biomedical Ontologies. In *American Medical Informatics Association Annual Symposium, AMIA'10*.
- Piasecki, M., Walkowiak, T., Rudnicka, E., Naskret, T., and Bond, F. (2017). The Concept of Lexical Platform. In *Proceedings of the LDK 2017 Workshops: 1st Workshop on the OntoLex Model (OntoLex-2017), Shared Task on Translation Inference Across Dictionaries & Challenges for Wordnets co-located with 1st Conference on Language, Data and Knowledge (LDK 2017), Galway, Ireland, June 18, 2017.*, pages 197–208.
- Schuler, K. K. (2006). *VerbNet: A Broad-Coverage, Comprehensive Verb Lexicon*. Ph.D. thesis, University of Pennsylvania.

- Teubert, W. (2007). *Text Corpora and Multilingual Lexicography*. John Benjamins Publishing Company.
- Urešová, Z., Fučíková, E., and Šindlerová, J. (2016). CzEngVallex: a Bilingual Czech-English Valency Lexicon. *The Prague Bulletin of Mathematical Linguistics*, 105:17–50.
- Urešová, Z., Fučíková, E., Hajičová, E., and Hajič, J. (2018). Creating a Verb Synonym Lexicon Based on a Parallel Corpus. In *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC'18)*, Miyazaki, Japan, May. European Language Resources Association (ELRA).
- Urešová, Z. (2011). *Valenční slovník Pražského závislostního korpusu (PDT-Vallex)*. Studies in Computational and Theoretical Linguistics. Ústav formální a aplikované lingvistiky, Praha, Czechia.
- Urešová, Z., Fučíková, E., Hajič, J., and Šindlerová, J. (2015). *CzEngVallex - Czech English Valency Lexicon*. Charles University in Prague, LINDAT/CLARIN PID: <http://hdl.handle.net/11234/1-1512>.

## 7. Language Resource References

- Cinková, S., Fučíková, E., Šindlerová, J., and Hajič, J. (2014). *EngVallex - English Valency Lexicon*. LINDAT/CLARIN PID: <http://hdl.handle.net/11858/00-097C-0000-0023-4337-2>.
- 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. (2012). *Prague Czech-English Dependency Treebank 2.0*, <https://catalog.ldc.upenn.edu/LDC2004T25>. Institute of Formal and Applied Linguistics, LINDAT/CLARIN PID: <http://hdl.handle.net/11858/00-097C-0000-0015-8DAF-4>.
- Hajič, J., Bejček, E., Bémová, A., Buráňová, E., Hajičová, E., Havelka, J., Homola, P., Kárník, J., Kettnerová, V., Klyueva, N., Kolářová, V., Kučová, L., Lopatková, M., Mikulová, M., Mírovský, J., Nedoluzhko, A., Pajas, P., Panevová, J., Poláková, L., Rysová, M., Sgall, P., Spoustová, J., Straňák, P., Synková, P., Ševčíková, M., Štěpánek, J., Urešová, Z., Vidová Hladká, B., Zeman, D., Zikánová, Š., and Žabokrtský, Z. (2018). *Prague Dependency Treebank 3.5*. Institute of Formal and Applied Linguistics, LINDAT/CLARIN, Charles University, LINDAT/CLARIN PID: <http://hdl.handle.net/11234/1-2621>.
- Hajič, J., Panevová, J., Hajičová, E., Sgall, P., Pajas, P., Štěpánek, J., Havelka, J., Mikulová, M., Žabokrtský, Z., Ševčíková, M., and Urešová, Z. (2006). *Prague Dependency Treebank 2.0*. LDC, number LDC2006T01, LINDAT/CLARIN PID: <http://hdl.handle.net/11858/00-097C-0000-0015-8DAF-4>.
- Pala, K., Čapek, T., Zajíčková, B., Bartůšková, D., Kulková, K., Hoffmannová, P., Bejček, E., Straňák, P., and Hajič, J. (2011). *Czech WordNet 1.9 PDT*. LINDAT/CLARIN PID: <http://hdl.handle.net/11858/00-097C-0000-0001-4880-3>.
- Urešová, Z., Štěpánek, J., Hajič, J., Panevová, J., and Mikulová, M. (2014). *PDT-Vallex: Czech Valency lexicon linked to treebanks*. LINDAT/CLARIN PID: <http://hdl.handle.net/11858/00-097C-0000-0023-4338-F>.