

MATEMATICKO-FYZIKÁLNÍ FAKULTA
PRAHA

**CzEngClass: Contextually-based Synonymy and Valency of Verbs
in a Bilingual Setting**

**(CzEngClass: Kontextová synonymie a valence sloves
v bilingvním prostředí)**

ZDEŇKA UREŠOVÁ, EVA FUČÍKOVÁ, EVA HAJIČOVÁ

ÚFAL Technical Report
TR-2019-62

ISSN 1214-5521



UNIVERSITAS CAROLINA PRAGENSIS

Copies of ÚFAL Technical Reports can be ordered from:

Institute of Formal and Applied Linguistics (ÚFAL MFF UK)

Faculty of Mathematics and Physics, Charles University

Malostranské nám. 25, CZ-11800 Prague 1

Czechia

or can be obtained via the Web: <http://ufal.mff.cuni.cz/techrep>

Annotation guidelines for CzEngClass lexicon

Abstract

The project focuses on contextually-based synonymy and valency of verbs in a bilingual setting. The analysis of semantic ‘equivalence’ (synonymy or near synonymy) of verb senses, and their valency behavior in parallel Czech-English language resources is the core of the proposed research. Using the translational context supports more language-independent specification of properties of verb sense classes of synonyms and leads towards generalization across languages. An initial sample bilingual verb lexicon of classes representing synonym or near-synonym pairs of verbs (verb senses) based on richly annotated corpora and existing lexical resources will be created. The main contribution of the project will be a deeper insight from the bilingual perspective into the topic of verb meaning in context based on the Functional Generative Description theory, thus extending it towards appropriate description of contextually-based verb synonymy.

Acknowledgements

This work described herein has been supported by the grant GA17-07313S of the Grant Agency of the Czech Republic, and it uses resources hosted by the LINDAT/CLARIN and LINDAT/CLARIAH-CZ Research Infrastructures, projects No. LM2015071 and LM2018101 supported by the Ministry of Education of the Czech Republic.

Contents

PART I – English version	4
1. Goal of the Annotation	5
2. Annotation Tasks	5
3. Annotation Tasks in the First Phase	5
3.1. How to Create a Preliminary Class of Verb Synonyms.....	5
4. Procedure for Selecting Class Members	9
5. Annotation Tasks in the Second Phase	9
5.1. How to Create Links to Other Lexical Sources	10
5.1.1. Links to OntoNotes	12
5.1.2. Links to VerbNet.....	13
5.1.3 Links to PropBank.....	16
5.1.4. Links to WordNet	17
5.1.5. Links to FrameNet	18
6. How to Create a Roleset	22
6.1. Repertoire of Semantic Roles in CzEngClass	23
6.2. Annotation of Semantic Roles in CzEngClass	25
7. How to Map Roleset Members to Validation Frame Members.....	29
7.1. The Semantic Role realization: Role Mapping	32
7.1.1. Mapping a Semantic Role to Alternating Functors	33
7.1.2. Mapping a Semantic Role to an Embedded Valency Frame Member	33
7.1.3. Mapping One Semantic Role to Several Arguments of One Valency Frame	34
7.2. The Functor's Specification when Mapping to a Semantic Role.....	36
8. Annotation Tools for Creating the CzEngClass Dictionary	42
8.1. Synonyms Editor	42
8.2. Editor Extension for Example Sentences Assignment.....	43
8.3. Editor Extension for English Verb Assignment to Class Name	43
9. Archiving and Distributing Annotated Data.....	44
PART II – Czech version	46
1. Cíl anotace.....	47
2. Úkoly anotace	47
3. Úkoly anotace v první fázi	47
3.1. Jak vytvořit předběžnou třídu slovesných synonym	47

4. Postup při výběru členů třídy	51
5. Úkoly anotace ve druhé fázi.....	51
5.1. Jak vytvořit odkazy do jiných lexikálních zdrojů	52
5.1.1. Odkazy do OntoNotes	54
5.1.2. Odkazy do VerbNetu	56
5.1.3 Odkazy do PropBanku	59
5.1.4. Odkazy do WordNetu.....	60
5.1.5. Odkazy do FrameNetu	62
6. Jak vytvořit Roleset	66
6.1. Repertoár sémantických rolí v CzEngClass.....	67
6.2. Anotace sémantických rolí v CzEngClass.....	69
7. Jak namapovat členy Rolesetu na členy valenčního rámce	73
7.1. K realizaci sémantické role: mapování rolí	76
7.1.1. Mapování sémantické role na alternující funktry	77
7.1.2. Mapování sémantické role realizované jinak než přímým valenčním členem	77
7.1.3. Mapování jedné sémantické role na několik argumentů jednoho valenčního rámce současně.	78
7.2. Specifikace funktoru při mapování na sémantickou roli.....	79
8. Anotační nástroje pro tvorbu slovníku CzEngClass	85
8.1. Synonyms Editor	85
8.2. Rozšíření editoru o přiřazování příkladových vět	86
8.3. Rozšíření editoru o přiřazování anglického slovesa do názvu třídy	86
9. Archivace a distribuce anotovaných dat	88
PART III – Technical guidelines.....	89
10. Installation of TrEd for Windows	89
10.1 How to install TrEd	89
10.2. Installation of additional Perl modules.....	91
10.3. How to Run TrEd and Install TrEd's Extensions	92
11. Installation of TortoiseSVN	95
12. Installation of CzEngClass Editor	96
12.1. Editor.....	96
12.2. Data	97
12.3. Configuration File Settings	98
12.4. How to run CzEngClass Editor	98
SVN Update and SVN Commit.....	99

PART I – English version

1. Goal of the Annotation

The annotation of the bilingual synonym lexicon is part of a three-year (2017-2019) project called *Contextually-based synonyms and valency of verbs in a bilingual setting*. This project is supported by the Grant Agency of the Czech Republic (No. GA17-07313S1).¹

The aim of the annotation is to create a lexicon of bilingual Czech-English synonyms. Synonyms are extracted from translated texts of the Prague Czech-English Dependency corpus. A functionally adequate relationship in terms of translation must exist between the meaning of the English and the Czech verb, i.e., the English and the Czech verb are synonymous in the given context(s) and the translated verb adequately expresses the functional intent of the original. We aim for each synonymous class to be characterized both meaning-wise (semantic roles) and structurally (valency arguments) by linking (mapping) semantic roles and valency members (Role_Argument mapping).

While valency frames are already specified for the verbs in the text and captured in the associated valency lexicons - PDT-Vallex (for Czech), EngVallex (for English) and CzEngVallex (for verbal translational equivalents), semantic roles and their association with arguments have not yet been processed.

2. Annotation Tasks

Several tasks need to be accomplished in order to meet the annotation objective. We have divided them into the two following phases:

- In the first phase, the goal of the annotation is to create **preliminary classes** of verb synonyms based on the translation from English to Czech.
- In the second (main) phase, the goal is
 - to create a "**Roleset**" (a set of semantic roles) for each class;
 - to link (map) semantic roles and valency members (**Role_Argument mapping**).
 - To provide **links** to other lexical (internal and external) resources.

3. Annotation Tasks in the First Phase

In the first phase, our goal is **to create preliminary classes of verb synonyms**. We base this task on texts translated from English to Czech contained in the Prague Czech-English Dependency Corpus (The PCEDT²).

3.1. How to Create a Preliminary Class of Verb Synonyms

¹ <https://ufal.mff.cuni.cz/czengclass>

² <https://ufal.mff.cuni.cz/pcedt2.0/en/index.html>

At this stage, the job of the annotator is to manually go through the automatically generated English verbal translational counterparts of the Czech verbs and create candidates for English verb members of one synonymous class. In this annotation we use excel files and apply the following criterion.

The English and the Czech verb must be in a functionally adequate translation relationship. This means that the translated verb adequately expresses the functional intent of the original verb. Existing PCEDT texts are used for determining the context of the translation.

An example of a functional translation of the verb *learn* using the verb *dozvědět se* in context:

the **PCEDT** Subsequently, **I_{ACT}** have **learned** that_{PAT} a private group, of which *Du Pont* is a part, **is_{PAT}** **funding_{PAT}** a modest program to continue data gathering at the Scott report stations as well as to develop more sophisticated UVB measuring instruments.
the **PCEDT** Následně **jsem se dozvěděl**, že_{PAT} nezávislá skupina, jejíž součástí je *Du Pont*, **financuje_{PAT}** malý program na pokračující sběr dat na stanicích uvedených ve Scottově zprávě a na vývoj výkonnějších přístrojů na měření UVB záření.

At the same time, it should be taken into account that a single verb can have multiple meanings that are - in the context of the PCEDT - determined by a valency frame. The valency frames are based on the valency theory as elaborated in the Functional Generative Description and captured in the respective valency lexicons: CzEngVallex³, PDT-Vallex⁴, and EngVallex⁵.

An example of a valency frame for one verb meaning of the verb *hradit* (*pay*) – in the PCEDT, the frame is marked with an identifier: *hradit* (*v-w1116f1*).

hradit¹ _{48x, 20x} **ACT(1) PAT(4) ?ADDR(3) ?EFF(zá+4)**

(platit) za byt jim hradil všechny poplatky; zaměstnavatel h. zaměstnancům pojištění; z Fondu.MEANS nár.majetku; za vzorné chování.CAUS jim h. výdaje na cestu; za děti.SUBS h. všechny výdaje rodiče

An example of a valency frame for one verb meaning of the verb *pay* – in the PCEDT, the frame is marked with an identifier: *pay* (*ev-w2223f7*).

pay⁶ **ACT() PAT() ?ADDR() ?EFF({for}[objpp;ving];to+inf)**

Investors pay higher prices for country funds.

Theatres pay movie producers for showing their films.

Mr. Cray was paid \$600,000 at Cray Research last year.

³ <https://lindat.mff.cuni.cz/services/CzEngVallex/>

⁴ <http://lindat.mff.cuni.cz/services/PDT-Vallex/>

⁵ <http://lindat.mff.cuni.cz/services/EngVallex/>

*The Latin American nation has paid very little on its debt.
...the TV commercial paid for *trace* by Giuliani's campaign...*

The excel File (Fig. 0) contains class and class member pairs. These pairs were automatically generated from the PCEDT translations. The file has 8 columns.

A	B	C	D	E	F	G	H
Class	Classmember	Status	Groupings	Sense	CzEngVallex	Restrictions	Notes
1	dát (v-w369f21)						
2							
3	dát PDT-Vallex-ID-v-w369f21						
4							
5							
6	put EngVallex-ID-ev-w2449f7_u_nobody	groupings (put)			czengvallex: put(ev-w2449f7_u_nobody) - dát(v-w369f21)		
7							
8	put EngVallex-ID-ev-w2449f15_u_nobody	groupings (put)			czengvallex: put(ev-w2449f15_u_nobody) - dát(v-w369f21)		
9							
10	put_up EngVallex-ID-ev-w2454f1	groupings (put_up)			czengvallex: put_up(ev-w2454f1) - dát(v-w369f21)		
11							
12	set EngVallex-ID-ev-w2900f8	groupings (set)			czengvallex: set(ev-w2900f8) - dát(v-w369f21)		
13							
14							
15							
16	datovat se (v-w11235hsa_989)						
17							
18	datovat se PDT-Vallex-ID-v-w11235hsa_989						
19							
20	date EngVallex-ID-ev-w808f2	groupings (date)			czengvallex: date(ev-w808f2) - datovat se(v-w11235hsa_989)		
21							
22							
23	dit se (v-w492f2)						
24							
25	dit se PDT-Vallex-ID-v-w492f2						
26							
27	be EngVallex-ID-ev-w218f3	groupings (be)			czengvallex: be(ev-w218f3) - dit se(v-w492f2)		
28							
29	happen EngVallex-ID-ev-w1548f5	groupings (happen)			czengvallex: happen(ev-w1548f5) - dit se(v-w492f2)		
30							
31							
32	dolahovat (v-w614f1)						

Fig. 0 Annotation of Class Status and Sense with Restrictions and Notes

In **column A (Class)**, the provisional name of the class is represented by the Czech verb sense and its valency frame identification, while each verb sense has its own identifier distinguishing verb meanings. For technical reasons, the identifier is a more complex entry and not just a numerical index (1, 2, 3) common in dictionaries. For example, *čekat* (*wait*) has one meaning with ID *v-w311f1*, and another meaning with ID *v-w311f9_ZU* in the PDT-Vallex.

In **column B (Class member)**, there are potential members of the given class (English verbs in one synonymous verb meaning).

For example, two potential English members are listed for the class *čekat* (*v-w311f9_ZU*): the first verb sense of the verb *wait* with EngVallex-ID-*ev-w3576f3* and the second sense of *wait* with EngVallex-ID-*-ev-w3576f4*.

In **column D (Groupings)** and in **column F (CzEngVallex)**, there are links to other sources to help the annotator decide whether or not to include the potential member (verb sense) into the class.

Annotations are written into columns C and E.

Column C (Status) is used to mark whether or not the verb sense is included in the class. For this annotation, we select one of the following values:

Y (Yes) - the verb sense belongs to the class,

N (No) - the verb sense doesn't belong to the class.

If unsure, we have the option to assign one of the following values; however, we try to use these values as little as possible:

R_Y (Rather Yes) - the verb sense is more likely to belong in the class,

R_N (Rather No) - the verb sense is more likely not to belong to the class.

We may also encounter a case where a verb has been assigned incorrectly to a given class by the automatic process. If we think that's the case, we assign the value:

D (Delete) - the verb sense does not belong to the class, it is an obvious error that occurred already in the automatic verb pairing process that took place in the data set from where the pairs were automatically preselected.

In **column E (Sense)**, we fill out the number of the English verb sense that corresponds to the meaning of the given class. A set of meanings for a given English verb sense can be found by clicking on a link in column D - an entry from OntoNotes (Groupings). We fill out column E only if we classify the verb sense as the class member of the given class. A member may have one or more meanings in OntoNotes. If we cannot find the appropriate meaning in OntoNotes, we write "NO" in column E.⁶

Column F (CzEngVallex) contains references to the CzEngVallex valency dictionary for both the Czech verb sense (specifies *Class*) and the English verb sense (specifies *Class member*). The link allows you to look at examples (sentences) in which both verbs were used. A click in the Excel file redirects you to the online version, where sample sentences can be loaded by clicking on *Show corpus examples*.

Column G (Restriction) is used for inserting a condition that a given verb must fulfill when it is a member of a given synonymous class. We record only free text explaining in what context the English verb can be considered synonymous with the Czech verb. There are no format requirements. But the important thing is not to use de-spacing.

Column H (Notes) is meant for notes. For the sake of automatic processing of annotated data it is important not to use multi-line annotations..

⁶ In OntoNotes verbs such as *build_up*, *look_for*, etc. are included with verbs without underscores - i.e., *build*, *look*. If the link in OntoNotes does not open the correct page with individual meanings, we look for the correct meaning on this page: http://verbs.colorado.edu/html_groupings/.

4. Procedure for Selecting Class Members

When selecting potential class members, we apply the following procedure, divided into 4 steps:

Step 0

An automatic selection of 200 Czech verbs (frames) - which provisionally denote class names

Step 1

An automatic selection of English verb translation counterparts of verbs from Step 0, class assignment (status annotation), general annotation of the individual class members, i.e., *Roleset*, mappings, links, examples, restrictions, notes

Step 2

An automatic selection of Czech counterparts of verbs from step 1 (except verbs from Step 0), class assignment (status annotation), general annotation of the individual class members, i.e., *Roleset*, mappings, links, examples, restrictions, notes

Step 3

An automatic selection of English verb translation counterparts of verbs from Step 2

If there are more verb members, further steps will follow (4, 5...).

5. Annotation Tasks in the Second Phase

In the second phase, the aim of the annotation is to create a “***Roleset***” for each class (Sec. 6) and **map** its members to the members of the valency frame of each class member (Sec. 7). We use references to other lexical sources (Sec. 5.1). These references are also subject to annotation.

Pre-prepared verbs in the individual classes must meet the following conditions:

1. All verbs of a given class must have the same set (*Roleset*) of semantic roles. An example: Recipient, Phenomenon, Source – the class *dozvědět se* (*learn, know, hear, doslychat se, naučit se, zaslechnout....*)
2. The English translational verb sense should have similar valency to the Czech verb in the given sense. An example: *dozvědět se*: ACT, PAT, ORIG = *learn*: ACT PAT ORIG

For this annotation phase we use a special annotation editor - **CzEngClass editor** (Sec. 8.1.), called **Synonyms Editor (SynEd)**. SynEd captures (**Fig. 1a**, from left to right) *Classes*, *Class Members* and *syntactic-semantic information (SynSem)*, *links (Links)* and *examples (Examples)*. The use of this editor is described below and in Section 8.

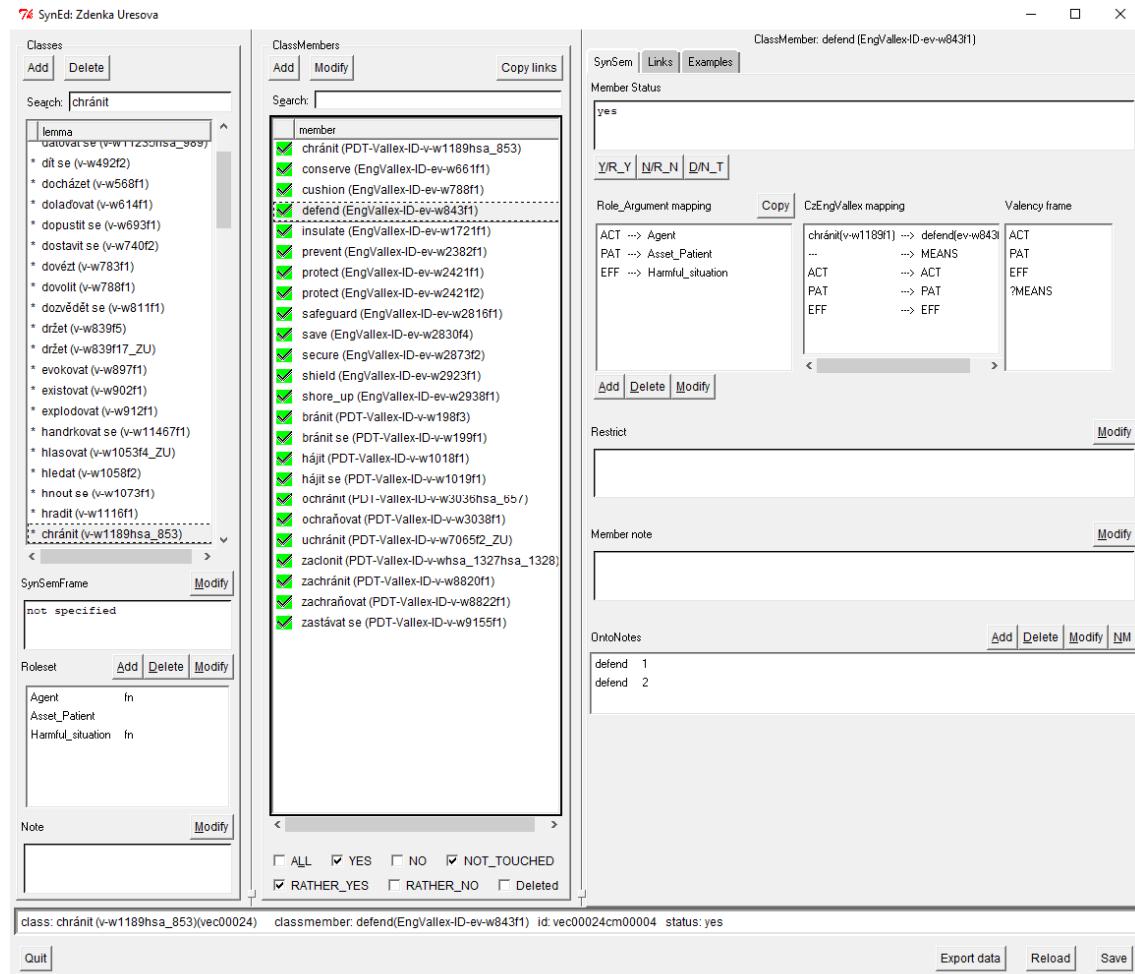


Fig. 1a CzEngClass editor (SynEd) - Display for annotator *Zdenka Uresova*, class *chránit* [protect] (*v-w1189hsa_853*), its member *defend* (*EngVallex-ID-ev-w843f1*)

5.1. How to Create Links to Other Lexical Sources

We want references to other lexical sources to be part of the CzEngClass dictionary for two reasons. We assume that linking our dictionary to other relevant lexical sources will allow it to be interoperable, and thus the dictionary will have a higher value for use in computer linguistics.

At the same time, links to other semantic databases give us a better view of the individual meanings of verbs and their characteristics. In addition to checking⁷ the pre-prepared links to PDT-Vallex, EngVallex, and CzEngVallex, the aim of the annotation is to assign the

⁷ References may be invalid due to changes in the dictionaries (PDT-Vallex, EngVallex, CzEngVallex) and due to the continuous annotation of the CzEngClass dictionary.

following links in SynEd, recorded in the **Links** tab, to individual Class Members (*CMS*) (Fig. 1b - Links⁸):

- for Czech CMSs: to Vallex,
- for English CMSs: to OntoNotes, FrameNet, WordNet, VerbNet and PropBank.

References (links) to individual sources can be multiple, but also one or none.

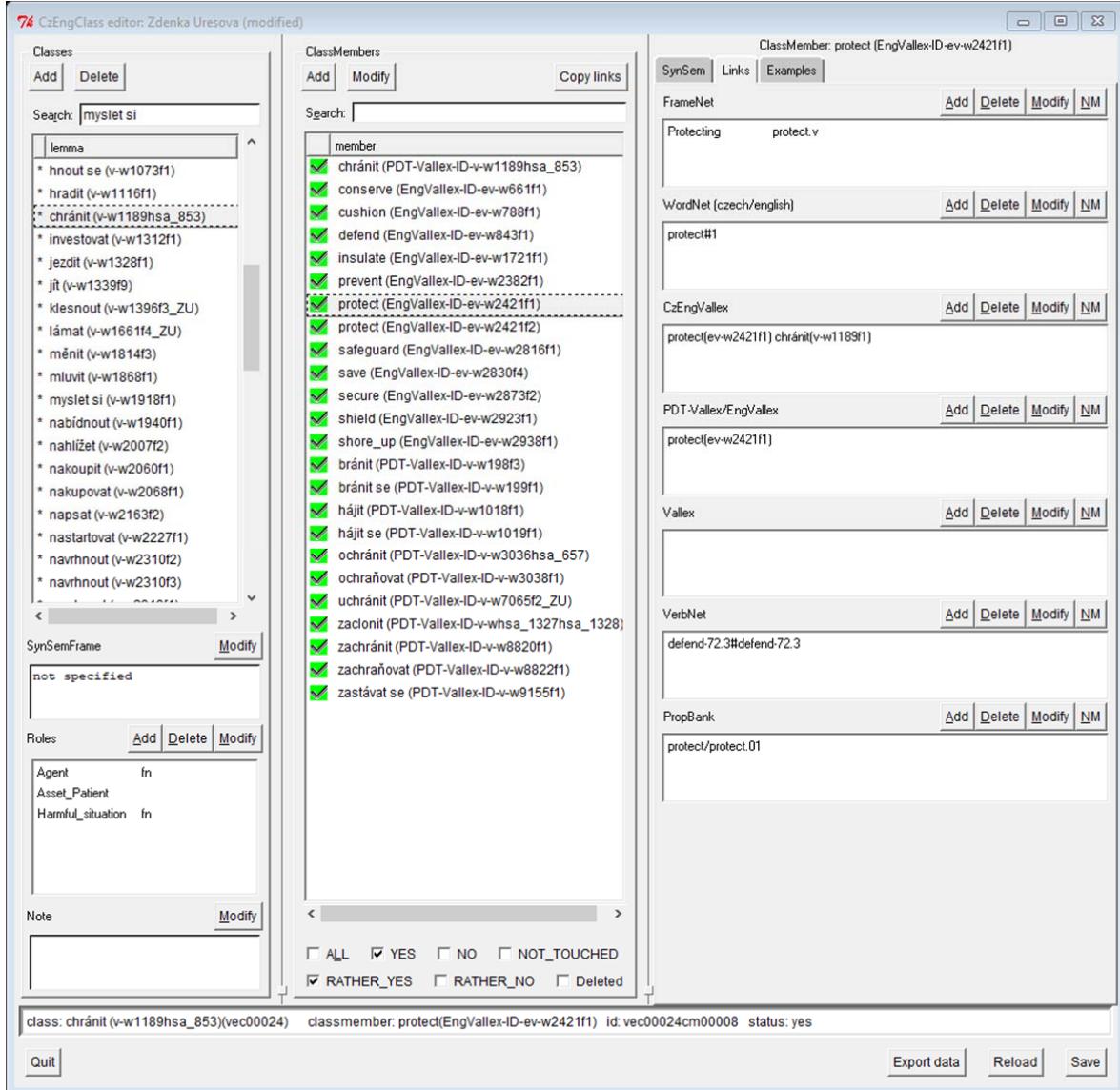


Fig. 1b. Showing links in the SynEd editor for the verb *protect* in the *chránit* class (NoA)

⁸ Due to the continuous capturing of the annotation rules, some images are no longer up to date. However, we keep them in their original form as they illustrate the process and evolution of annotations. We always point out the inaccurate spots in the manual by the phrase ‘not up-to-date – NoA’.

5.1.1. Links to OntoNotes

A reference to OntoNotes/Groupings is the key reference for the English verb. In *SynEd*, we open the website http://verbs.colorado.edu/html_groupings/defend-v.html with a link to OntoNotes by double-clicking the appropriate English CM. Based on the context, i.e., the examples in the PCEDT (so far – NoA - in *SynEd* made available in TrEd via the *Examples* tab > Show in TrEd tab), the correct Sense is determined (Fig. 2).

Sense Number 2: protect or speak up against criticism or opposition

Commentary: Syntax Is: NP1[agent[human]] DEFEND NP2[topic[abstract]]

NOTE: generally implies a verbal endorsement of a process, state or concept

NOTE: includes legal courtroom defense.

Examples:

John defended his thesis brilliantly before the committee.

The administration is defending its economic policies.

He hasn't agreed to defend our position at tomorrow's meeting.

I won't defend her actions in this matter.

Bob defended his little sister against their mother's criticisms.

They are fighting to defend their rights.

He vigorously defended his point of view.

That law firm has a reputation for defending members of the mob.

The famous attorney will defend her in the high-profile murder case.

Mappings:

VerbNet: defend-72.2

FrameNet: Justifying

PropBank: defend.01

WordNet 3.0 Sense Numbers: 1, 5, 6, 7

Fig. 2 Selected sense from OntoNotes for the verb *defend*

For each meaning recorded in OntoNotes under *Mappings*, we find references to the given verb sense to VerbNet, FrameNet, PropBank, and WordNet (e.g., for the verb *defend*, sense 2):

Mappings:

VerbNet: defend-72.2

FrameNet: Justifying

PropBank: defend.01

WordNet 3.0 Sense Numbers: 1, 5, 6, 7

These links listed in OntoNotes *Mappings* can be used as the default for the links' annotation in the CzEngClass dictionary. All references must be checked, since they might no longer be

valid (due to changes in the annotation versions in the linked resources or due to other reasons). Displaying links to OntoNotes in SynEd is captured in **Fig. 3** (verb *protect - sense 1*, bottom right). There can be zero links (record as *No Mappings - NM*), only one link, or more links. Use the *Add* button to add links. Like other items in the editor, you can also modify (*Modify*) or delete (*Delete*) it.

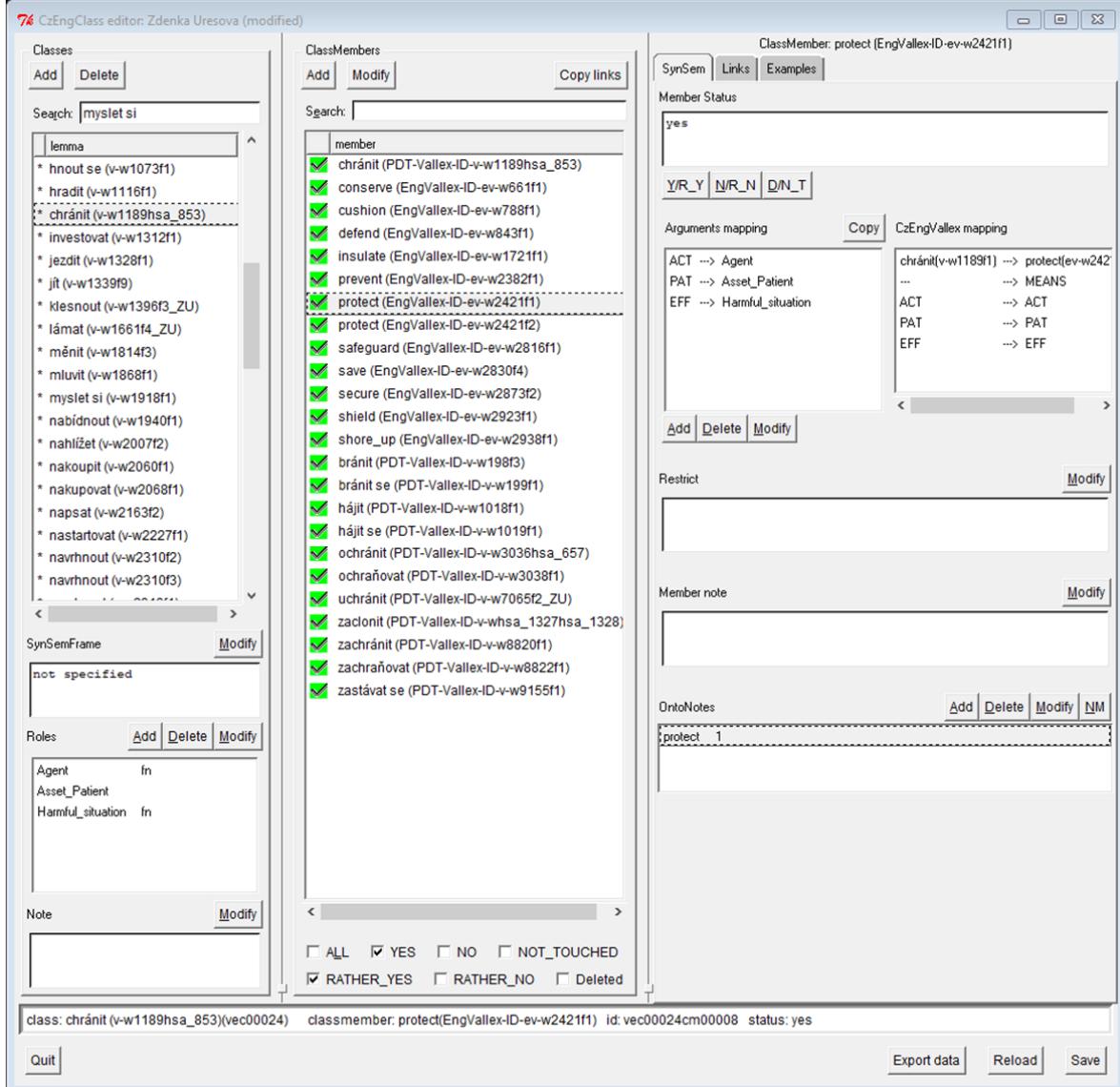


Fig. 3 Depiction of links of CM *protect* to OntoNotes in SynEd ((NoA)

5.1.2. Links to VerbNet

It is useful to continue the annotation of links by assigning links to VerbNet. If we know a VerbNet link for a given CM from the *Mappings* listed in OntoNotes, we use the *Add* button

to assign the link and via the *Show* button (**Fig. 4**) we check if the link is correct. The *Show* button takes us to the annotation of the given meaning of the verb in VerbNet.⁹

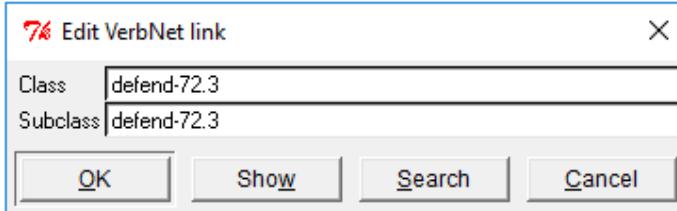


Fig. 4 Editing links to VerbNet for the verb *defend*

If the *Show* button doesn't work, we use the *Search* button (**Fig. 4**), which takes us to the Semlink page.¹⁰ Here, we find a given verb (*defend*)¹¹ via the *Search* button (Fig. 5).

Fig. 5 Annotation for the verb *defend* in SemLink

By clicking on the *defend-72.3* class, we get more annotation details: **DEFEND (wn 1, 6, 7; g 1, 2, 3)**, that we can utilize. The abbreviations in brackets after the verb are links to FrameNet (missing here), WordNet (**wn 1,-6,-7**) and Groupings (**g 1,-2,-3**). You can click to see the link.

⁹ <http://verbs.colorado.edu/verb-index/vn3.3/vn/defend-72.3.php#defend-72.3>

¹⁰ <http://verbs.colorado.edu/verb-index/vn3.3/>

¹¹ <http://verbs.colorado.edu/verb-index/vn3.3/search>

If an OntoNotes link does not exist for a given CM, or if we cannot select one (NM is listed in *SynEd* for both cases), using the *Add* button (**Fig. 6**) we open the Unified Verb Index¹², where we use *Search* to search for a given verb. A picture for the verb *defend* verb is shown in **Fig. 5**, for the verb *save* in **Fig. 7**.



Fig. 6 The window for assigning links to VerbNet - Edit buttons top right

 A screenshot of a web browser displaying search results for the verb "SAVE". The URL is "verbs.colorado.edu/verb-index/vn3.3/search.php". The search term is "SAVE". The results are organized into several sections:

- VERBNET MEMBERS:** SAVE: **BILL-54.5**, SAVE: **GET-13.5.1-1**
- VERBNET CLASSES:** NO VERBNET CLASS MATCHES
- VERBNET THEMROLES:** NO THEMROLES MATCHES
- VERBNET PREDICATES:** NO PREDICATES MATCHES
- PROPBANK:** SAVE: **SAVE,V**
- FRAMENET:** NO FRAMENET MATCHES
- ONTONOTES SENSE GROUPINGS:** SAVE: **SAVE,V**, SAVE.N: **SAVE,N,V**

 At the bottom, it says "This page generated on 2017.8.15 at 2:46 PM." and provides links to "REFERENCE", "CLASS HIERARCHY", "CONTACT", "INSPECTOR", "VxC", and "GENERATOR".

Fig. 7 An image for the searched verb *save* in SemLink

On the left side of the screen below *VerbNet Members*, or *VerbNet Classes*, we find the displayed annotations of the searched verb in SemLink (e.g., *save* in class *bill.54.5* and *get-13.5.1-1*).

If no class from VerbNet fits our annotated CM, we record the appropriate verb (as in the case of the verb *save*, as CM of the class *protect*, Fig. 7) link status to Verbnet as *No Mapping* - via *NM* button (Fig. 6).

If we can include the searched verb in one of VerbNet's red-listed classes, we click on the selected class (*VerbNet Members*, or *VerbNet Classes*, as appropriate) (**Fig. 8** for the verb *defend*) and insert its name as a link via the *Add* button (**Fig. 6**) to *SynEd* (the class name can be copied for convenience). We can also check whether the references listed in VerbNet (for

¹² <http://verbs.colorado.edu/verb-index/vn3.3/>

the verb *defend*: **DEFEND (WN 1, 6, 7; G 1, 2, 3)** match our annotation of links. If so, we click the links and use them to insert more links into *SynEd*.

The screenshot shows the VerbNet v3.3 interface. At the top, there are links for 'RETURN HOME | BACK | SEARCH' and 'VIEW OR MANAGE ALL COMMENTS | UNIVERSITY OF COLORADO'. The main title is 'defend-72.3' with 'Members: 8, Frames: 4'. Below this, there is a 'No Comments' section and a 'POST COMMENT' button. To the right, there is a 'CLASS HIERARCHY' section showing 'DEFEND-72.3' with 'NO SUBCLASSES'. A 'KEY' button is located in the top right corner. The 'MEMBERS' section lists the following words with their WN and G codes:

- DEFEND (WN 1, 6, 7; G 1, 2, 3)
- GUARD (WN 1, 2; G 1)
- INSURE (WN 3; G 1)
- PRESERVE (WN 2; G 1)
- PROTECT (WN 1; G 1)
- SAFEGUARD (WN 1; G 1)
- SHADE (WN 3)
- SPEAK_FOR (WN 1; G 4)

Fig. 8 Annotation of the verb *defend* in class *defend-72.3*

5.1.3 Links to PropBank

To assign the link(s) to PropBank, you can use the help from the *Mappings* listed in OntoNotes (Fig. 2 *defend*): **PropBank: defend.01**. We will assign the reference **defend.01** in a similar way as the reference to VerbNet, but we will use the *PropBank* window (Fig. 9) and write the window ID in the *Roleset* accurately, e.g. 01, not 1 (Fig. 10).



Fig. 9 Link Assignment Window to PropBank - Edit buttons top right



Fig. 10 Editing links to PropBank for the verb *defend*

To check the annotation we open the annotation pages of the respective verb in PropBank through the *Show* button.¹³

¹³ <http://verbs.colorado.edu/propbank/framesets-english-aliases/defend.html>

We can also utilize the references in SemLink (**Fig. 5**). Here the annotation of the respective verb is captured in the right section under PropBank.

If there is no help for our annotation task, we open - again via the *Search* button (**Fig. 10**) - the SemLink website,¹⁴ where we look up the corresponding verb and its annotation in PropBank. Here, the verb is captured even if it is not annotated in VerbNet (**Fig. 11** – for the verb *cushion*).

The screenshot shows a search interface for the verb 'cushion'. At the top, there are links to 'RETURN HOME', 'BACK', and 'SEARCH'. The search bar contains the query '[cUSHION]'. Below the search bar, there are four main sections: 'VERBNET MEMBERS' (No matches), 'PROPBANK' (CUSHION: CUSHION.V), 'VERBNET CLASSES' (No matches), and 'FRAMENET' (No matches). On the left, there are additional sections: 'VERBNET THEMROLES' (No themroles matches) and 'ONTONOTES SENSE GROUPINGS' (CUSHION: CUSHION.V). At the bottom, there are links to 'REFERENCE', 'CLASS HIERARCHY', 'CONTACT', 'INSPECTOR', 'VxC', and 'GENERATOR'. A note at the bottom left states 'This page generated on 2017.8.15 at 2:46 PM.'

Fig. 11 Link for the verb *cushion* to PropBank via the SemLink database

5.1.4. Links to WordNet

To assign the link(s) to English WordNet, you can again use the help from the *Mappings* listed in OntoNotes (**Fig. 2** *defend*): WordNet 3.0 Sense Numbers: 1, 5, 6, 7. Those links have to be added to the *WordNet* window (**Fig. 12**),

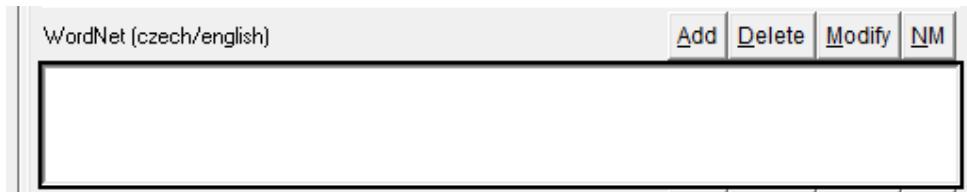


Fig. 12 WordNet Link Assignment Window - Edit buttons top right

where we record in the *Sense* window (**Fig. 13**) the number of the appropriate meaning of the annotated verb corresponding to the meaning given in the WordNet database (**Fig. 14**).¹⁵ Only the number (3), not a number with a cross (#3) should be recorded into *SynEd*.

¹⁴ <http://verbs.colorado.edu/verb-index/vn3.3/>

¹⁵ <http://wordnetweb.princeton.edu/perl/webwn?&o7=1&s=defend>

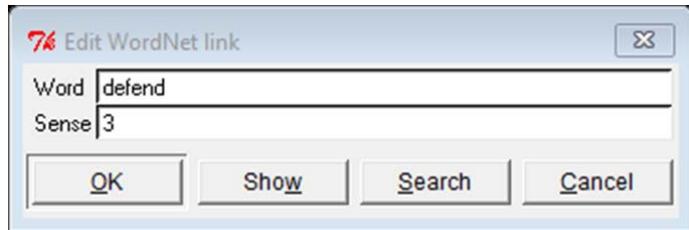


Fig. 13 Editing links to WordNet for the verb *defend*

WordNet Search - 3.1

- [WordNet home page](#) - [Glossary](#) - [Help](#)

Word to search for:

Display Options:

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations
 Display options for sense: (gloss) "an example sentence"
 Display options for word: word#sense number

Verb

- S: (v) defend#1, support#8, fend for#1 (argue or speak in defense of) "She supported the motion to strike"
- S: (v) defend#2 (be on the defensive; act against an attack)
- S: (v) defend#3, guard#3, hold#20 (protect against a challenge or attack) "Hold that position behind the trees!"; "Hold the bridge against the enemy's attacks"
- S: (v) fight#2, oppose#2, fight back#2, fight down#1, defend#4 (fight against or resist strongly) "The senator said he would oppose the bill"; "Don't fight it!"
- S: (v) champion#1, defend#5 (protect or fight for as a champion)
- S: (v) defend#6, represent#8 (be the defense counsel for someone in a trial) "Ms. Smith will represent the defendant"
- S: (v) maintain#8, defend#7 (state or assert) "He maintained his innocence"

Fig. 14 WordNet database search for the verb *defend* meanings

If we don't have a hint for WordNet links, we look in the WordNet database, which we simply open with the *Search* button in the referencing window to WordNet (**Fig. 12**).¹⁶

5.1.5. Links to FrameNet

From the point of view of semantic roles (SRs), the most important source and inspiration for our annotation is the FrameNet electronic database¹⁷, which works with semantic roles (frame elements - FEs) within semantic frames. We can also be inspired by VerbNet's¹⁸ theta roles.

¹⁶ <http://wordnetweb.princeton.edu/perl/webwn>

By making a link to FrameNet's semantic frame(s) for the annotated CM, we can consider to use the SRs assigned to that frame. FrameNet link(s) are recorded in the *FrameNet* window (**Fig. 15**).



Fig. 15 FrameNet Link Assignment Window - Edit buttons top right

To assign the FrameNet links we follow a similar approach as described above. *FrameNet* Editing Window (**Fig. 16**) includes a window for the FrameNet semantic frame name (*Frame name*) and a window for the lexical unit name to refer to (*LU name*).

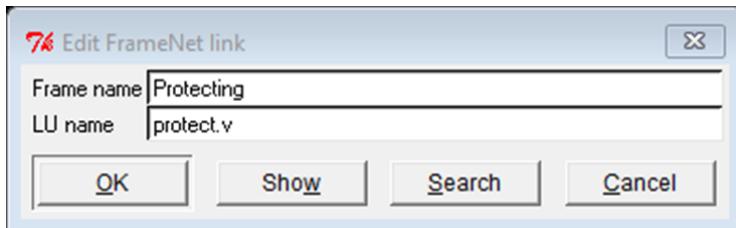


Fig. 16 FrameNet Editing Window

If we have a link for the annotated CM (*protect*, *chránit* class) - thanks to the mappings indicated for the given verb in OntoNotes - we will just record it in the *FrameNet* editing window and verify its validity in the *Show* button (**Fig. 16**). The name of the semantic frame (*Protecting*) is listed in the *FrameNet* database in the top right-hand corner (**Fig. 17**).

¹⁷ <https://framenet.icsi.berkeley.edu/fndrupal/about>

¹⁸ <https://verbs.colorado.edu/~mpalmer/projects/verbnet.html#thetaroles>

Annotation Lexical Entry Protecting

protect.v

Frame Element	Core Type
Asset	Core
Beneficiary	Extra-Thematic
Danger	Core
Degree	Extra-Thematic
Depictive	Extra-Thematic
Descriptor	Extra-Thematic
Duration	Extra-Thematic
Event description	Extra-Thematic
Explanation	Extra-Thematic
Instrument	Peripheral
Manner	Peripheral
Means	Peripheral
Place	Peripheral
Protection	Core
Time	Peripheral

[Turn Colors Off](#)

- Wenemy,forces,defense-(1)
 1. The lumbering carts and chariots are PROTECTED by outriders mounted on giant wolves who patrol the area in front of the advancing tribe , probing for enemies and scouting for small settlements to loot [CNI]
 2. If the water table is within 15 m of the surface , the cheapest method is to dig a well by hand and PROTECT the walls from collapse by building a masonry lining [CNI]
 3. It was an old graveyard , PROTECTED by the high walls from the bustle and intrusion of the outside world
- NP-T-NP-PPfrom-(1)
 1. These regulations are designed to PROTECT people against immediate or delayed health risks from substances which may have an adverse effect on their health ; for instance , disinfectants such as glutaraldehyde
 2. IN THE MENDIPS A SHEEP FARMER IS STRUGGLING TO PROTECT HIS LAMBS FROM THE FOXES THAT WATCH AND PREY .
 3. Paint PROTECTS your car from rust
 4. Iod Richardson PROTECTS himself from burglars .
 5. But China is also PROTECTING North Korea from the wrath of the United Nations Security Council .
- NP-T-PPfrom-(1)
 1. Children must be PROTECTED from indoctrination from such sources [CNI]
 2. Scottish parents are apparently even more likely to do this -- one in three parents never PROTECT their children from its harmful rays despite the fact that many of them have

Fig. 17 FrameNet reference for the verb *protect*, class *chránit*

Click the name of a semantic frame to open its definition in FrameNet (**Fig. 18**). Alongside the definition, the appropriate semantic roles (Frame Elements - FEs) are summarized.

<https://framenet2.icsi.berkeley.edu/fnReports/data/frameIndex.xml?frame=Protecting&banner=1>

Frame Index

Protecting

[Lexical Unit Index](#)

Definition:

Some **Protection** prevents a **Danger** from harming an **Asset**.
The rubber on her sneakers PROTECTED her from the mud.

The coral reefs **SHIELDED** the coastline from the worst of the waves.

A growing number of sustainable farmers are preserving agricultural variety and **PROTECTING** biodiversity by raising ?heritage? or ?heirloom? varieties.

The man responsible for this entire ordeal **cowardly SHIELDED** himself from justice with the little boy's struggling body.

FEs:

Core:

- Asset [Asset]**: Something desirable possessed by or directly associated with the **Protection** which might be lost or damaged.
Almost 10 years later, that agreement **INSULATED** Microsoft from liability on copyright claims over the graphical user interface (GUI).
- Danger [Danger]**: A situation that could damage the **Asset**.
Other nations tended to muffle or modulate their occasional annoyance with the country that was their **SHIELD** against the Soviet threat.
- Protection [pro]**: The person, entity, or action that prevents harm to an **Asset**.
The tabloid's deal with the woman **SHIELDED** Schwarzenegger.

Non-Core:

- Beneficiary [I]**: FN: This extra-thematic FE applies to participants that derive a benefit from the protecting event.
- Degree [I]**: This frame element selects some gradable attribute and modifies the expected value for it.
- Depictive [I]**: FN: This FE describes a participant of the state of affairs introduced by the target as being in some state during the action.
- Descriptor [des]**: A description or characteristic of the **Protection**.
- Duration [I]**: FN: This FE denotes the length of time in which the **Asset** is protected.
- Event description [eve]**: This FE is used for phrases that describe the protecting event as a whole. It applies to appraisals of the event, typically a judgment of its expectedness. The notion of appraisal requires a factive portrayal of the event. It also applies to descriptions of the protecting event as a role filler in another frame.

Fig. 18. Definition and semantic roles for FrameNet's semantic frame *Protecting*

If we agree with the given reference, the annotation is done. But if we do not have a reference for a given CM, or disagree with it, we proceed as follows.

Using the *Search* button in the FrameNet editing window in *SynEd* (Fig. 16) we open the FrameNet lexical unit index,¹⁹ where we enter the CM into the tab search (Fig. 19). If the CM is included in FrameNet, we find the frame that best suits our annotation (for the verb *protect*, we select *Protecting*) in the search list of assigned semantic frames. Sometimes several compliant frames can be referenced to the CM. If that is the case, we record them all in the editing window via the *Add* button, or we use the *OK + Next* buttons. If we don't find an appropriate link, or none fits, we record the *NM* status.

¹⁹ <https://framenet.icsi.berkeley.edu/fndrupal/luIndex>

The screenshot shows a web browser window with the URL <https://framenet.icsi.berkeley.edu/fndrupal/lulIndex>. The page title is "LU index". At the top, there's a navigation bar with links: "About FrameNet", "Documentation", "FrameNet Data", "Related Projects", and "Bibliography". Below the navigation bar is a search bar containing the word "protect". To the right of the search bar is a "Search" button. Below the search bar, there's a horizontal menu with letters A through Z and a "All" option. The main content area is titled "FrameNet Index of Lexical Units" and contains a sub-section titled "Search: protect". This section lists three items under a bullet point:

- protect.v ([Protecting](#)) Finished_Initial [Lexical Entry Annotation](#)
- protection [entity]n ([Protecting](#)) Created [Lexical Entry Annotation](#)
- protection [event]n ([Protecting](#)) Created [Lexical Entry Annotation](#)

Fig. 19 Searching the FrameNet database

6. How to Create a Roleset

The **Roleset** captures the meaning of the class of Czech and English synonymous verbs; it **is defined by a set of Semantic roles** (SRs) that is common to all class members (CMs) of one synonymous class. We create this set of roles by linking (mapping) semantic roles with valency members of the valency frame for the individual CMs that are recorded in a synonymous class (Sec. 7).

Roleset examples for the class:

"*klesnout-decline*": Item, Initial_value, Final_value, Difference

"*nabídnout-offer*": Offerer, Recipient, Entity_offered, Entity_received

"*chránit-defend*": Agent, Asset_Patient, Harmful_situation

The Roleset's annotation is very closely related to the mapping of valency members to the individual semantic roles (Sec. 6). This is because the possibility of mapping the valency behavior of a verb sense in a given class to the semantic roles determines the Roleset.

Deciding which roles to select for a given synonymous class is not a simple matter that would be possible right after the annotation of one class member. Often, it is only during the annotation of all potential class members that the semantic character of each class member becomes clear to us based on the already annotated links, and it also becomes clear which SRs have to be assigned to the arguments from the valency frame of each CM.

Each synonym class has its **Typical Mapping of Roleset and Valency Frame Arguments (TypRAM)** that corresponds to the majority mappings in the given class. At the same time,

the typical mapping corresponds to the mapping given for both Czech and English verbs selected for the class name (see Sec. 8.1.3. for class name annotation).

TypRAM examples for the class:

klesnout (v-w1396f3_ZU) – *decline* (ev-w829f2):

ACT > *Item*,
ORIG > *Initial_value*,
PAT > *Final_value*,
DIFF > *Difference*

nabídnout (v-w1189hsa_853) – *offer* (ev-w2105f1):

ACT > *Offerer*,
ADDR > *Recipient*,
PAT > *Entity_offered*,
EFF > *Entity_received*

chránit (v-w1189hsa_853) – *defend* (ev-w843f1):

ACT > *Agent*,
PAT > *Asset_Patient*,
EFF > *Harmful_situation*

6.1. Repertoire of Semantic Roles in CzEngClass

The exact repertoire of semantic roles in the CzEngClass dictionary has not yet been determined; its production is also part of the annotation. There are different approaches to the semantic roles inventory, which varies in size by approach: from large (about 1,300 FEs in FrameNet), to medium (36 theta roles in VerbNet) and to small (2 proto-roles in the Dowty 91 concept, or 2 macroroles in the Van Valin 99 concept²⁰).

We assume that the repertoire of roles in CzEngClass (**CzEngClass-Roles-Inventory**, **CERI**), given the purpose for which we use them (synonymous bilingual classes), will not literally replicate any existing concept created for one language. Nonetheless, given the longstanding tradition of working with SRs, we have no ambition to create entirely new role labels, but we are inspired by labels already established.

The starting point for **CERI** is the annotation of SRs in FrameNet, mainly because they are created for specific semantic frames that can relate to meaning-related verbs. But we plan to differ significantly from FrameNet in the number of roles included in CERI. We try to keep the number of roles to a minimum, but we also want them to convey meaning. We expect to

²⁰ Dowty, David R. 1991. Thematic proto-roles and argument selection. *Language* 67:547–619

Van Valin, Robert D., Jr. 1999. Generalized semantic roles and the syntax-semantics interface. In *Empirical issues in formal syntax and semantics*. Vol. 2. Edited by Francis Corblin, Carmen Dobrovie-Sorin, and Jean-Marie Marandin, 373–389. The Hague: Thesis

have a repertoire of about 200 SRs. In the current version, CERI includes (**Fig. 22**) 89 SRs taken from FrameNet (**Tab. 1**) and 23 newly created SRs (**Tab. 2**).

Act	Interlocutor 2
Action	Issue
Activity	Item
Addressee	Label
Agent	Manner
Agreement	Medium
Appraisal	Message
Area	Money
Arguer 1	Money owner
Arguer 2	Offerer
Assessor	Opinion
Asset	Origin
Authority	Participant 1
Buyer	Participant 2
Cognizer	Patient
Cognizer agent	Perceiver
Comparator	Perceiver agentive
Complainier	Perpetrator
Complaint	Phenomenon
Components	Prize
Content	Profiled item
Created entity	Project
Creator	Proposal
Difference	Proposition
Driver	Protagonist
Earner	Purpose
Earnings	Recipient
Entity	Requirement
Seller	Support
Sender	Supported
Signatory	Supporter
Sought entity	Target currency
Source	Theme
Source currency	Topic
Speaker	Undesirable situation
Standard item	Value
State	
Stimulus	

Tab. 1 List of SRs taken from FrameNet (current version in SynEd)

Activity Event Process	Driver Entity
Agent Entity	Entity offered

Area 1	Entity received
Area 2	Goal Destination
Asset_Patient	Means Vehicle
Asset_Theme	Patient_Entity
Audience_Addressee	Phenomenon_Fact
Benefactive	Proposition_Fact_Phenomenon
Benefactor	Recipient Place
Cognizer_Inspector	Speaker_Agent
Cognizer_Item	Speaker_Medium
Delivered_entity	Topic_Message
	Transporter

Tab. 2 List of newly created SRs, NOT taken from FrameNet (current version in SynEd)

Given that we are still building new synonym classes, both the rules for creating new roles and the existing list of roles (**CERI**) are yet to be finalized. We realize that annotating more verbs may affect not only the list, but also the rules for creation of SRs.

6.2. Annotation of Semantic Roles in CzEngClass

The annotation of SRs in the CzEngClass dictionary means that we assign SRs to the Roleset (**Fig. 20**) of individual synonymous classes, thereby creating also the repertoire of SRs in the CzEngClass dictionary (CERI).

When annotating SRs and including them into the Roleset, we distinguish two situations:

Situation 1:

If we have a link to the semantic frame in FrameNet for a given CM at our disposal, we take SRs from that semantic frame and include them in CERI.

When editing these roles in *SynEd*,

- we either select the SRs from a list of pre-prepared roles (**Fig. 21** and **Tab. 1**),
- or we add them to the list (**Fig. 22 – Add role**, in defining the role we remember to click *From FrameNet* – **Fig. 23a**).

Situation 2:

If we do not have a link to the semantic frame in FrameNet for the given CM at our disposal, we either draw on the roles featured in FrameNet for related verbs,

When editing these roles in *SynEd*,

- we either select the SRs from a list of pre-prepared roles (**Fig. 21** and **Tab. 1**),
- or we add them to the list (**Fig. 22 – Add role**, in defining the role we remember to click *From FrameNet* – **Fig. 23a**).

or we create our own labels for SRs.

When editing these roles in *SynEd*,

- we either select the SRs from a list of pre-prepared roles (**Fig. 21 and Tab. 2**),²¹
- or we add them to the list (**Fig. 22 – Add role**, in defining the role we remember to click *NO From FrameNet* – **Fig. 23b**).

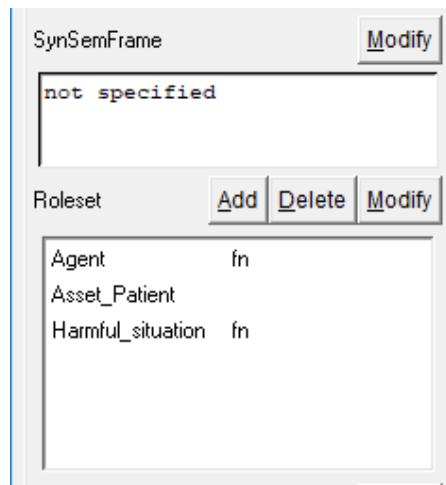


Fig. 20 The *Roleset* Window in *SynEd*

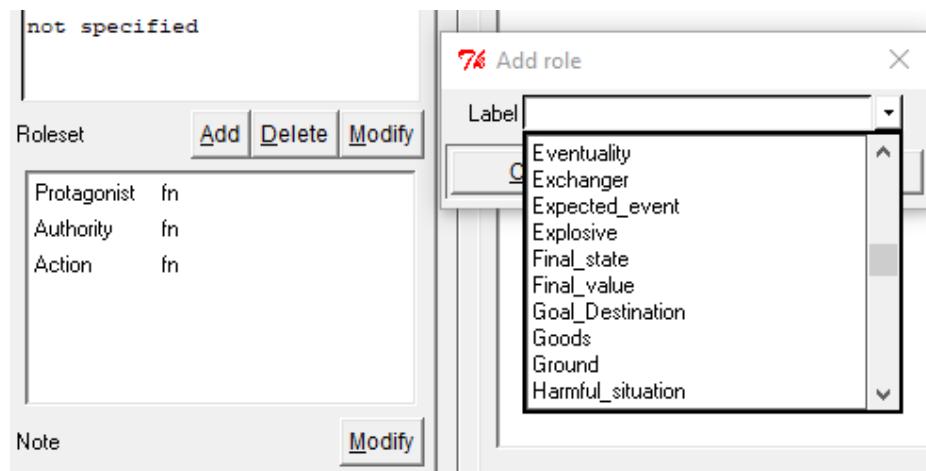


Fig. 21 Selecting SRs from CERI

²¹ In the *Add Role* window, the Role List does not distinguish which role is from FrameNet and which is not. Role origin is indicated only in the *Roleset* window (the *fn* marks the roles from FrameNet) on the right side of the role.

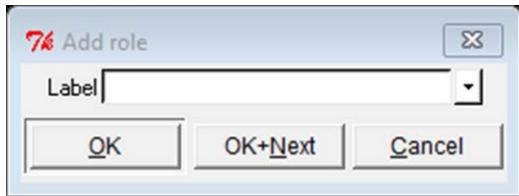


Fig. 22 Adding a new SR

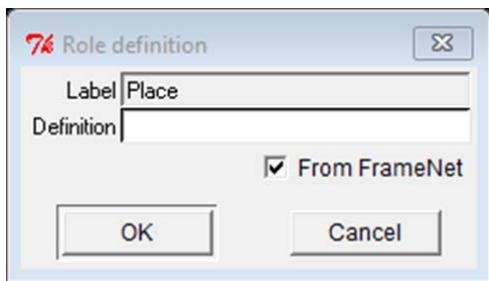


Fig. 23a Definition of a new role- **taken** from FrameNet (check *From FrameNet*)

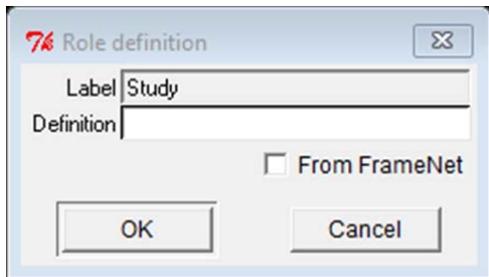


Fig. 23b Definition of a new role- **not taken** from FrameNet (do not check *From FrameNet*)

We do not fill in the definition of a semantic role. This is included later.

Situation 1 - we have a link to the FrameNet semantic frame for the given CM at our disposal

If multiple CMs of the synonymous X-class contain a reference to a common or similar semantic frame in FrameNet, the annotation of SRs in Roleset is simpler. We check the repertoire of FEs in these common frames and select from the list of FEs (Fig. 21) those SRs (or FEs) that best fit our annotation. It must be taken into account that FEs, as representatives of semantic information within a given semantic frame, are seen not as universal semantic roles, but as specific to the frame. Therefore, for example, with CMs class *protect*, we find several possible SRs for the mapping of the Value Actor (ACT), depending on which semantic frame each CMs in FrameNet refers to.

Examples:

Class member *save* (*ev-w2830f4*) refers to the semantic frame of *Rescuing* where the SR *agent* matches the actor:

<https://framenet2.icsi.berkeley.edu/fnReports/data/frameIndex.xml?frame=Rescuing&banner=>

Rescuing

An Agent saves a Patient or an Asset from a Harmful situation.

Israeli commandos RESCUED the hostages and left.

-Class member *defend* (*ev-w843f1*) refers to the semantic frame of *Defending* where the actor is matched by the SR *Defender*:

<https://framenet2.icsi.berkeley.edu/fnReports/data/frameIndex.xml?frame=Defending&banner=>

Defending

A Defender responds to an Assailant's attack on a Victim by directly, and usually violently, countering it.

Samantha stalwartly DEFENDED the gateway from the wraiths, crouching low over Stephen's prone form.

Class member *protect* (*ev-w2421f1*) refers to the semantic frame *Protecting* where the actor is matched by the SR *Protection*:

<https://framenet2.icsi.berkeley.edu/fnReports/data/frameIndex.xml?frame=Protecting&banner=>

Some Protection prevents a Danger from harming an Asset.
The rubber on her sneakers PROTECTED her from the mud.

Based on FEs comparison, we select the most appropriate "label" for the semantic role mapped to the valency member. For the above example of a *chránit/protect* class, we consider *Agent* the most appropriate label of the semantic role mapped to the actor.

If we do not find a role in CERI, we fill it in the list (Fig. 22 – *Add role*, when defining role²² we do remember to check *From FrameNet* – Fig. 23a).

²² We do not annotate the semantic role definition yet.

There are two ways to annotate the *Roleset* for a given synonymous class:

1. Either we record the individual SRs of the *Roleset* directly in the window of Roleset (Fig. 20) in *SynEd* by using the edit button *Add*,
2. or we first insert the SRs when annotating the SRs mapping to valency members for individual CMs in the *Role_Argument mapping* window (Fig. 25). See Sec. 7 for further information about editing in this window.

Situation 2 – we don't have a link to FrameNet semantic frame for a the given CM at our disposal

If none of the CMs of the synonymous class contain a reference to the semantic frame in FrameNet (this happens exceptionally), we can draw on the roles listed in FrameNet for related verbs. If we do not find a suitable role in the FrameNet list (Fig. 21 and Tab. 1), we either add a role to this list (Fig. 22 – *Add* role, in defining the role we remember to click *From FrameNet* – Fig. 23a), or we create our own label. With the *Add* button, we open a window for adding a role (Fig. 22), and after the *Role definition* window opens, we don't check *From FrameNet* (Fig. 23b).

The new label for a semantic role should adequately capture the content of a given participant, similar to FEs.²³ While roles from FrameNet are generally made to fit one semantic frame, they do not go across all frames (e.g., the role *Goal* expressing a purpose within one frame, can express a place within another frame), in the CzEngClass dictionary we tend to make SRs have the same meaning in all the classes in which they are used. The new label can be one-word (*Transporter*) or even multi-word (*Activity_Event_Process*). If we create a multi-word SR label, we write the label with the underscore (X_Y_Z). If words from multi-word naming could exist as separate role labels (SR: *Topic_Message* x SR: *Topic*, SR: *Message*), we write these words with a capital letter. This type of multiword label is used for a role that expresses both (or more) roles from the given label (*he said of Paul. Topic that he's lying. Message vs. he said that Paul was lying. Topic_Message*).

If a word within a multi-word label of SR specifies another word within this multi-word label, we use lowercase for that word (SR: *Entity_offered*, SR: *Entity_received*).

7. How to Map Roleset Members to Validation Frame Members

Each member of a synonymous class must meet the definition for inclusion in the relevant synonymous class. By this definition, only those translational verbs (verb meanings) that have similar meaning and structure characteristics are included in one synonymous class. In other words, the translational verbs of one synonymous class have the same or similar, i.e. synonymous, meaning and their valency frames are also the same or similar.

With regard to the annotation this means that for each CM of a synonymous class X, the semantic roles (*Roleset*) contained in *Roleset* (Sec. 6) need to be meaningfully mapped to valency members (*Arguments*) of the valency frame of this verb as a class member of class X.

²³ When creating new role labes, we can also be inspired by the list of thematic roles in VerbNet:
https://verbs.colorado.edu/verb-index/VerbNet_Guidelines.pdf.

Ideally, all CMs of a given class would have only one mapping. That may not be the case, therefore, we work with ***Typical Roleset Mapping and Valency Frame Arguments*** (**TypRAM**), which corresponds to the majority mapping in a given class (Sec. 6) and characterizes the meaning and structure characteristics of the chosen synonymous class.

When mapping the roles from a given *Roleset*, each of the roles must be mapped to "SOMETHING" from the valency frame of a verb in that class (see below an example of the class *vyhodnotit/evaluate*). The "SOMETHING" may be

- either a member of the valency frame (in the class *vyhodnotit/evaluate* functors in black)
- or any other free modification to which the given SR might have been mapped (in the class *vyhodnotit /evaluate* functors in blue)
- or a proxy functor (#any, #sb, #sth), (in the class *vyhodnotit /evaluate* functors in green)

Conversely, each member of the valency frame of a verb listed in the given synonymous class must be mapped to a semantic role from the assigned *Roleset*.

There is only one exception to this rule: If the valency frame of an English Class member²⁴ includes a non-obligatory free modification, written as a free modification with a question mark, we may not (but can) take it into account in the mapping when we create a *Roleset*, i.e., we may not (but can) take care of the mapping for this valency frame member.

If any member of the valency frame of a potential class member of the given synonymous class REALLY cannot be mapped to the chosen *Roleset* of that class, that potential class member (candidate) does not belong to the class.

Example:

Mapping for *vyhodnotit /evaluate* class

Roleset			
	<i>Assessor</i>	<i>Phenomenon</i>	<i>Value</i>
<i>Vyhodnotit</i>	ACT	PAT	COMPL
<i>Evaluate</i>	ACT	PAT	COMPL
<i>Rate</i>	ACT	PAT	EFF
<i>Dostat</i>	ORIG	ACT	PAT
<i>Ocenit</i>	ACT	PAT	EFF
<i>Posoudit</i>	ACT	PAT	COMPL
<i>Vyčíslovat</i>	ACT	PAT	EFF
<i>Zhodnotit</i>	ACT	PAT	#sth

A note about *Roleset* is written in the *SynEd* in the *Note* window below the *Roleset* window.

²⁴ It only applies to English verbs, because when EngVallex was created there were modified rules for the content of the valence frame: the frame can include non-obligatory free modifications.

How to choose a proxy functor (#any, #sb, #sth) or a specific functor listed in the valency theory of FGP.

If we are able to use a functor listed in the valency theory of FGD for the given valency member that we need to map, we prefer it to a proxy functor.

If we are not able to use such a listed functor, but we still need to map the valency member because it can be visualized for the given verb, we add one of the following proxy functors: **#sb (animate)**, or **#sth (unanimate)**, or **#any (both)** to the mapping.

In *SynEd*, we annotate this mapping under the Syntactic-Semantic Information tab (*SynSem*) in the *Role_Argument mapping* window (**Fig. 25**) as follows.

First, we open the *SynSem* tab (**Fig. 24**) in *SynEd*,

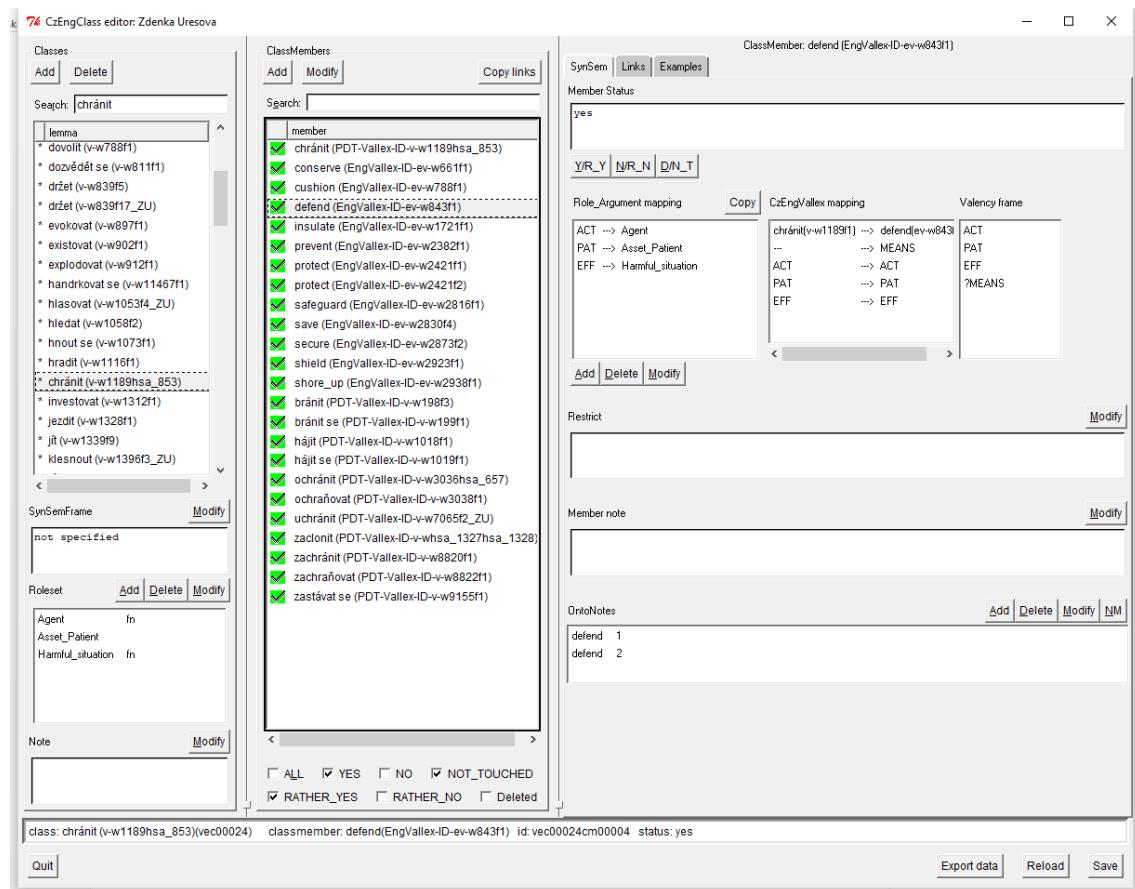


Fig. 24 *SynSem* tab with windows: *Role_Argument mapping*, *CzEngVallex mapping* and *Valency frame* for CM *defend* in the class *chránit/protect* (NeA)

where the following information is already displayed about the valency frame (VF) for each class member (*defend*) of the synonymous class (*chránit/protect*) - *Valency frame* window (ACT, PAT, EFF, ?MEANS) and the mapping of this VF to the VF of the Czech class

member verb (*chránit*) according to which the class is named - *CzEngVallex mapping* window (---> MEANS, ACT > ACT, PAT > PAT, EFF > EFF).

The *CzEngVallex mapping* and *Valency frame* window are not the editing windows, the information in them is intended to help to annotate the *role - argument mapping*, that is supposed to be written in the SynSem tab on left: *Role Argument mapping* (Fig. 25).

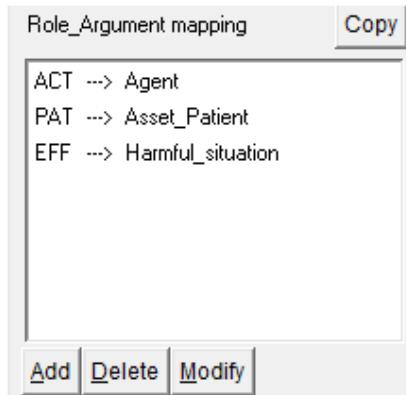


Fig. 25 The *Role_Argument mapping* window (CM *defend* class *chránit/protect*) - edit buttons lower left and copy mapping button upper right

If we have already prepared a *Roleset* for the class, the situation is simple. Use *Add* to open the mapping window (Fig. 26) and gradually map the valency members of the CM's of the synonymous class to the prepared roles via the *OK+Next* button.

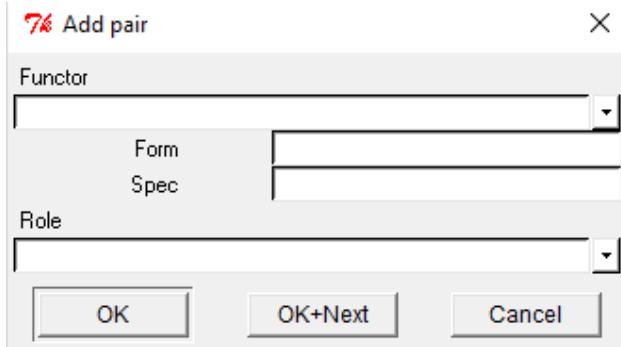


Fig. 26 The editing window for mapping

7.1. The Semantic Role realization: Role Mapping

In most cases, one SR is mapped to one argument from the valency frame of a class member.

One of the following situations may occur.

1. One SR is realized as a single argument but captured in the valency frame as several alternating functors, see Sec. 7.1.1.
2. One SR is realized in two (or more) ways, either (directly) as an argument of the valency frame of a class member or as a so-called "embedded" argument, see Sec. 7.1.2. By embedded argument, we mean an argument of a verb that is the root of a dependent clause representing some direct argument from a class member's valency frame.
3. One SR is realized by several arguments from one valency frame of a class member, Sec. 7.1.3.

7.1.1. Mapping a Semantic Role to Alternating Functors

To indicate semantic role mapping to alternating functors, select the proxy functor `#alt` in the *Functor* box (**Fig. 26**) and record all the alternating functors captured in the valency frame in the box *Spec* (**Fig. 26**). We separate them with a comma (NOTE: we don't use square brackets here).

Example:

The class *držet/hold* (*v-w839f17_ZU*) includes as a class member the verb *udržet* with a valency frame containing the alternating functor LOC and MANN.

udržet⁵ 3x,23x ACT(1) PAT(4) LOC(=)|MANN(=)

udržíme byt pohromadě.MANN(=); u. je naživu.MANN(=); u. ho ve vzorném pořádku.LOC(=); *udržet* děti v tichu.LOC(=); u. elektrárnu v provozu.LOC(=); u. syna v nevědomosti.LOC(=)

When mapping alternating functors to the semantic role, we select the proxy functor `#alt` from the menu of the functors and enter the alternating functors "LOC, MANN" in the *Spec* box. The following mapping will then appear in *SynEd*:

`#alt [LOC, MANN]->STATE`

If there were several alternating functors, we would write them all in the *Spec* box. For example: MANN, ACMP, CPR.

7.1.2. Mapping a Semantic Role to an Embedded Valency Frame Member

For the entry of *semantic role-argument mapping* for an SR that is realized directly as an argument of the valency frame of the class member we use the *editing window for mapping* (**Fig. 26**) and we map the SR to the corresponding valency member from the frame of the class member of the given synonymous class.

For the entry of *semantic role-argument mapping* for an SR that is **realized as an argument that is embedded in another “direct” argument** from a class member’s valency frame, we follow the following instructions:

The entry is formalized as follows.

1. Example: The entry **FUNC1|FUNC2** indicates that the role will be mapped to an embedded argument with FUNC2 depending on the valency argument of the verb with FUNC1. This entry is written in the *Spec* field without parentheses, added automatically by *SynEd*.
2. The entry **FUNC1|FUNC2|FUNC3]]** indicates that the role is mapped to an embedded argument with FUNC3 that is dependent on FUNC2 and FUNC2 depends on the valency argument of a verb with FUNC1. To enter **FUNC1|FUNC2|FUNC3]]** in the *Spec* field, brackets must be used for **FUNC2|FUNC3]**, as a result, the editor then adds one more bracket.

Example:

The **PAT[ACT]** notation for the class member *allow* (*ev-w86f4*) within the class *dovolit* (*v-w788f1*) indicates that the *Protagonist* role is an "embedded" **ACT** or that it depends on **PATient** of the verb *allow*.

The Food and Drug Administration.ACT allowed the company.ACT to begin.PAT arketting a new lens for use in cataract patients.

Example:

The **ACT|RSTR[PAT]]** notation for the class member *be* (*ev-w218f3*) within the class *dít* (*v-w492f2*) indicates that the *Entity* role is an "embedded" **PATient** that depends on the RSTR that depends on the ACT of the verb *be*.

There is nothing.ACT wrong.RSTR with the economy.PAT.

7.1.3. Mapping One Semantic Role to Several Arguments of One Valency Frame

One semantic role can also be realized by **several arguments** from one valency frame of a class member simultaneously - **structural splitting of one semantic role** into two valency frame members.

For example, the Roleset of the *stěžovat si* (*v-w652f1*) class contains three semantic roles: *Complainier*, *Addressee*, and *Complaint*. However, some members in this class have four arguments in the valency frame - ACT, ADDR, PAT, and EFF. Here, one semantic role (*Complaint*) is realized by two arguments (PAT and EFF) at once:

The following mapping applies for the realization of one role (**Complaint**) by two arguments
(PAT and EFF)

Výbor.ACT si jim.ADDR stěžuje na obchody.PAT, že neplní.EFF svou povinnost.

ACT > Complainier
ADDR > Addressee
PAT > Complaint
EFF > Complaint

To write a **mapping of one semantic role to one argument**, we do not need special instructions, as the straightforward realization of a semantic role corresponds to a straightforward mapping to one argument.

The verb *complain* from the *stěžovat si* (*v-w6521fl*) class has - unlike others class members - only three arguments in its valency frame - ACT, ADDR, PAT.

For the **realization of one role with one argument** the following mapping applies:

Foreigners.ACT complain to Washington.ADDR that they have.PAT limited access.

ACT > Complainier
ADDR > Addressee
PAT > Complaint

Structural splitting of one semantic role is typical for verbs of speaking, for example:

The class *soudit* (*v-w6270fl*) has two semantic roles in its Roleset: *Cognizer* and *Phenomenon*.

Some members of this class, e.g., the verb *soudit*, have ACT, PAT, and EFF in the valency frame. The mapping then maps **one role to two arguments**:

O exportu.PAT se soudí, že v srpnu silně vzroste.EFF

ACT > Cognizer
PAT > Phenomenon
EFF > Phenomenon

In the same class one role (Phenomenon) can be realized as only one argument.

Nicméně pokud ministr.ACT usoudí, že transakce učiní.PAT dopravce neschopným provozu.

ACT > Cognizer
PAT > Phenomenon

7.2. The Functor's Specification when Mapping to a Semantic Role

Sometimes the *functor* mapped to a *semantic role* needs to be further specified. We enter the specification in the *Spec* box (**Fig. 26**); the specification does not have a formally defined form yet, therefore use free notation but preferably unambiguous. Do not write square brackets when filling an entry to the *Spec* box. They are completed automatically.

Example:

The verb *vzepřít se* (*oppose*) has an identical valency frame (with ACT and PAT functors) for two meanings.

1. For the meaning of *opposing something inanimate* (*oppose her will, oppose his orders*) in the class *handrkovat se* (*haggle*) the Roleset contains: **Arguer1, Arguer2, and Issue**. The mapping should capture the possessive expression of the semantic role Arguer2.

```
ACT -> Arguer1  
#sb[poss.pronoun/adj.] -> Arguer2  
PAT -> Issue
```

2. The meaning of *opposing something animate* (*oppose the enemy*) does not require further specification, the Roleset: Arguer1, Arguer2 is mapped without any specification

```
ACT -> Arguer1  
PAT -> Arguer2
```

The valency member is represented by a *Functor* that has to be chosen from the pre-prepared list of functors by clicking on the arrow on the left (**Fig. 27**).



Fig. 27 The Functor's Annotation

The menu always features first those functors that are included in the valency frame of the verb (for *defend*: ACT, PAT, EFF, and MEANS). They are followed by the proxy functors

(#any, #sb, #sth) The absence of a valency member in the valency frame can have various reasons – an error in the frame, entry convention²⁵, etc.

An example of such mapping is the mapping of *Harmful_situation* role to the #sth proxy functor with the verb *conserve* class *chránit* (*protect*) (Fig. 28), whose valency frame does not include *EFF*, which is mapped to the role *Harmful_situation* with most class members of the class *chránit* (*protect*).

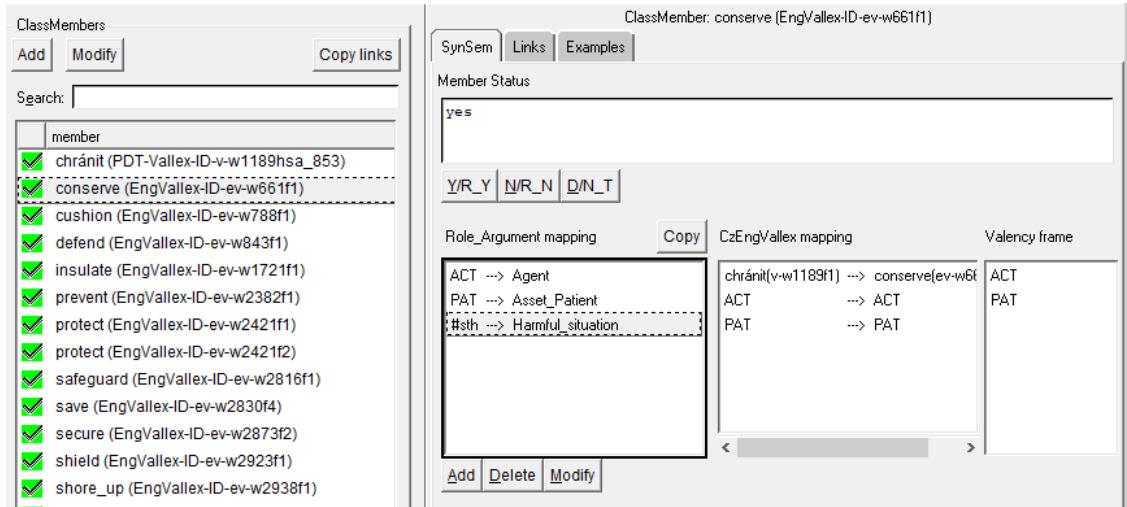


Fig. 28 Role mapping to a proxy functor for conserve

We enter the semantic role in the *Role* box; we select the semantic role from the SRs pre-prepared menu by clicking an arrow at the end of the *Role* box (Fig. 28).

²⁵ In the valency theory of the Functionally Generative Description (FGD), only obligatory (ACT, PAT, ADDR, EFF, ORIG) and optional (question mark before function, e.g., ?ACT) actants and obligatory free modifications (LOC, DIR) are entered in the valency frame. This holds for PDT-Vallex; in EngVallex, also non obligatory free modifications are written in the valency frame.

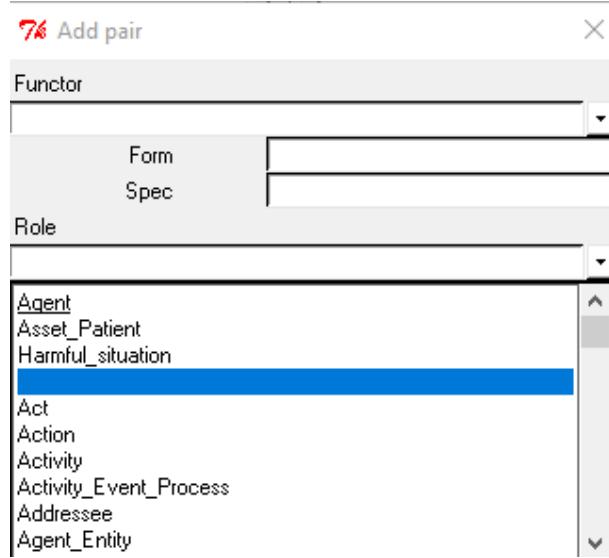
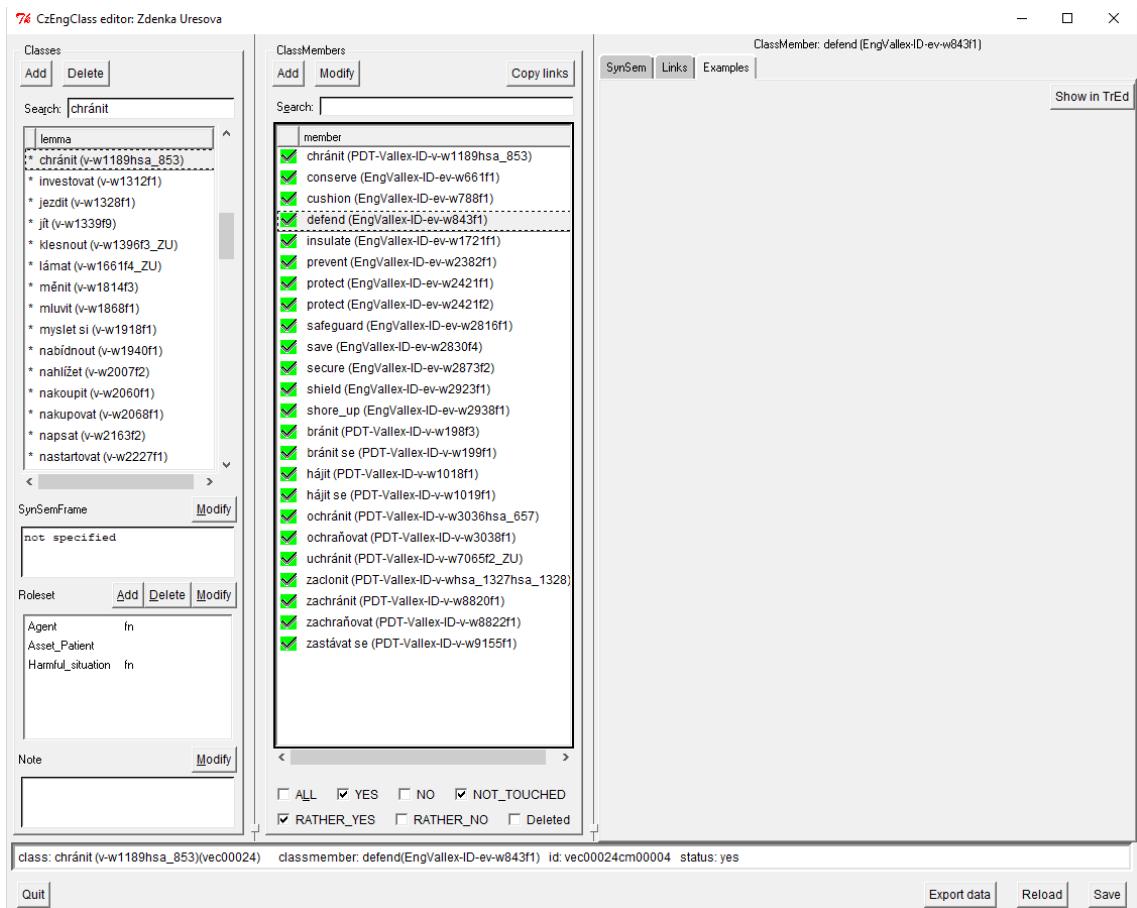


Fig. 28 The Semantic Role's Annotation

In the menu of Roles, the roles included in the Roleset of the verb appear first (for *defend*: Agent, Asset_Patient, Harmful_situation).

But if the *Roleset* for a given synonymous class has not yet been created (Sec. 6.2), we proceed as follows. By looking through FrameNet links for all CMs in a given class, we can see which FrameNet's semantic frame appears most frequently in the links, and examine its FEs, especially *Core FEs*. If other semantic frames for individual CMs in FrameNet links are captured, we check FEs for them as well. We prepare a list of them and consider which ones we select as members of our Roleset. If the labels offered by FrameNet's SRs (FEs) don't suit us, we either modify the labels or invent new ones. The newly created Roleset will then be mapped onto the valency frame's arguments of the individual CMs of the class, as described above.

Examples from the PCEDT can help us in mapping roles and arguments, as well as in annotation of a verbs' membership in a synonymous class (*Member Status*). They are implemented in the *SynEd* editor now but we can view them also by opening the translational pairs of verbs in the annotation tool *TrEd* (*Tree Editor*) in the *Examples* tab via the *Show in TrEd* button (Fig. 29 and 30).



Pic: 29 SynEd - The Example tab with the Show in TrEd button

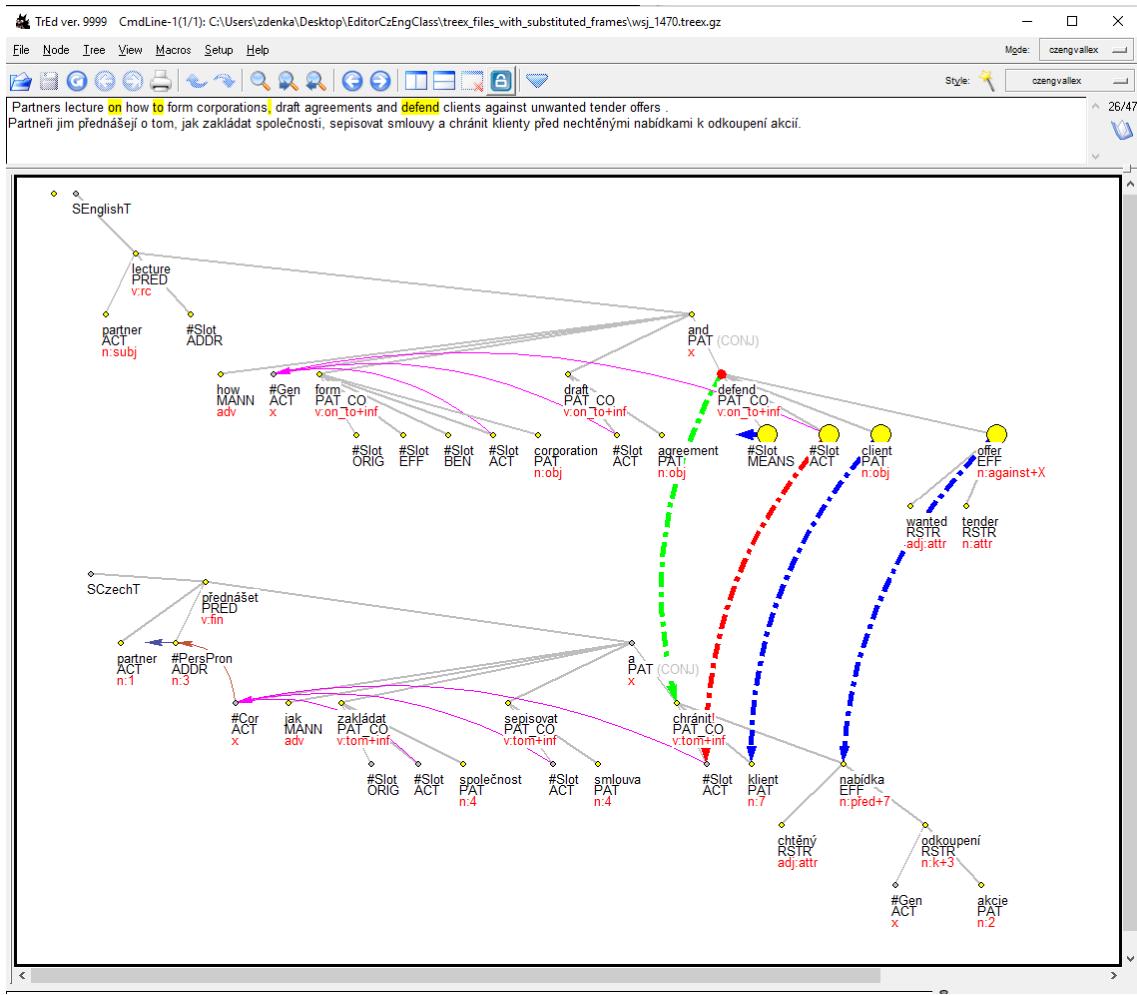


Fig. 30 Example for the translational pair *defend* – *chánit* in the PCEDT annotation, shown in TrEd editor

To facilitate the annotation, copy buttons are implemented in *SynEd* (Fig. 31 and 33). They can be used if the references (Fig. 31) or role mapping to arguments (Fig. 33) for one or more CMs of the class match.

To copy the links entered for one CM you must press the *Copy links* button (Fig. 31) when you “are” in the *ClassMembers* window on a CM whose links you want to copy, and confirm (Fig. 32) that you really want to do the copying. The *Links* are automatically copied to all CMs with the same lemma in the class (the class *chránit* – when you press *Copy Links* being on the CM *protect-f1*, the links are automatically copied to *protect-f2*).

Fig. 31 Copying links from one CM to another CM of the same class

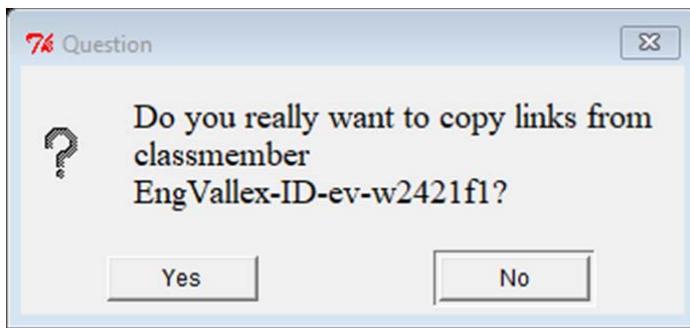


Fig. 32 Confirm Copy Links

We proceed similar if the role – argument mapping matches for one or more CMs of the class, except if we want to copy the mapping written for one CM, we press the *Copy* button (**Fig. 33**), which turns red; then we click on the CM where we want to copy the mapping and confirm that we want to proceed this operation.

If we copy the role – argument mapping to CM, where the mapping is already filled in, the editor will offer the *Merge* and *Replace* option. The *Replace* option will replace the existing mapping on the new CM with the mapping we want to copy, while the *Merge* option will add only those pairs from the copied source mapping (*Functor* → *Role*) that the existing mapping in the new CM did not include yet.



Fig. 33 Copying role - argument mapping from one CM to another CM of the same class

8. Annotation Tools for Creating the CzEngClass Dictionary

When creating the CzEngClass dictionary, we use two basic annotation formats, **Excel** (Sec. 3.1) and a special annotation editor *SynEd* (Sec. 8.1.). While we use Excel in particular for annotating the status of class members (class assignment), the *SynEd* editor is intended for general annotations.

8.1. Synonyms Editor

The annotation tool for the CzEngClass dictionary, called the ***Synonyms Editor*** (*SynEd*), was created "from scratch" during the first year of the project. The SynEd tries to make the work of the annotator as simple as possible and allows also for the transfer of the annotated information into the online form of a dictionary that we plan to make available online in the final year of the project as part of the services in the LINDAT/CLARIN repository.²⁶

The final version of *SynEd* has not yet been created. Some things have not yet been implemented, others are changed according to the demands of the annotators. Even so, we hope that the editor helps you to annotate the CzEngClass dictionary and make your effort easier. We will be grateful for any comments on the editor.

The **SynEd installation instructions** are part of the “technical” manual that is available to you. Instructions on how to use *SynEd* are described above. When annotating, it is important

²⁶ <https://lindat.mff.cuni.cz/cs/>

to continuously *save* and *commit* your work to the SVN system (Sec. 9) so that the annotations are not lost.

8.2. Editor Extension for Example Sentences Assignment

New example sentences as well as the *Add to Lexicon*, *Remove from Lexicon*, *Show one in TrEd*, *Show all in TrEd* buttons are added in the editor (**Fig. 34**). We select and add the most representative sentences (max. 5) to the dictionary.

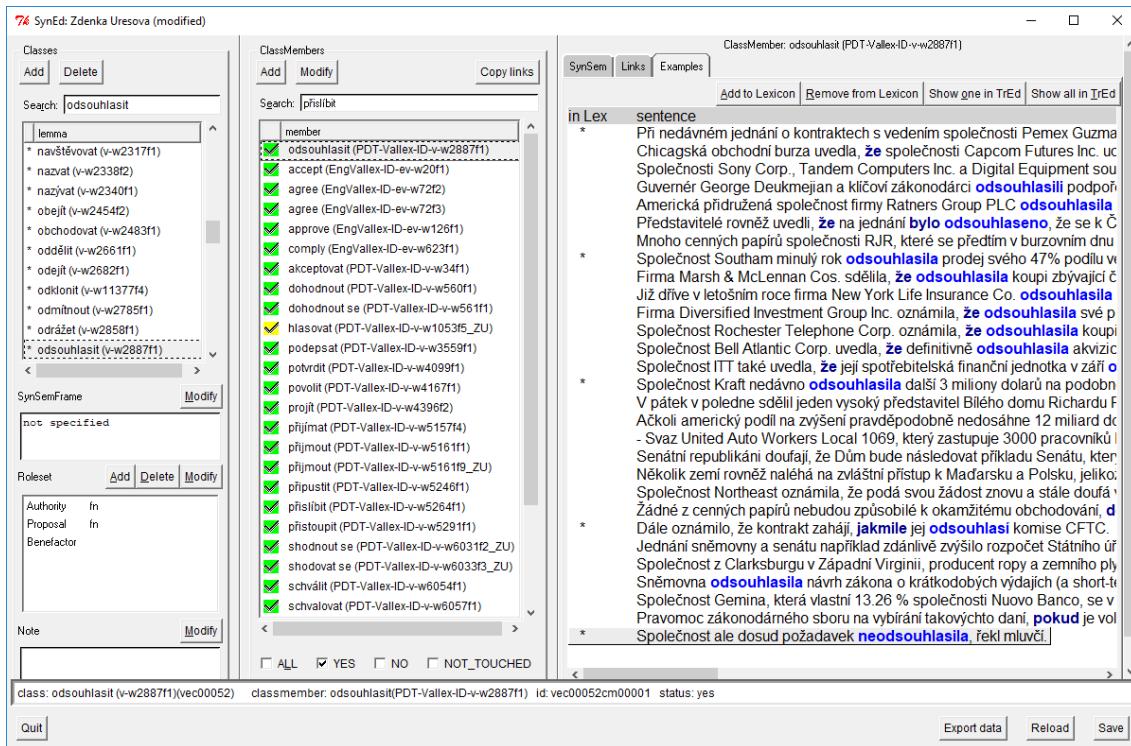


Fig. 34 *SynEd* - The *Example* tab with *Add to Lexicon*, *Remove from Lexicon*, *Show one in TrEd*, *Show all in TrEd* buttons

8.3. Editor Extension for English Verb Assignment to Class Name

For a long time the name of the synonymous class was only provisional and was represented by the Czech verb and the identification of the valency frame (Sec. 3.1.) - e.g., the verb *prozkoumat* (*v-w4633f1*) (*explore*).

The new editor extension allows assigning also a representative English verb to the class name. While the Czech verb in the class name is not annotated in any way (it is automatically preselected), the English verb has to be selected by the annotator from the assigned English

class members. We proceed as follows: we click the *Set* button in the *English Class Name* window (**Fig. 35**) and the English verb we have selected is automatically entered here.

We consider a representative English member of the class to be an English verb that has *Typical Mapping of Roleset and Valency Frame Arguments* (**TypRAM**) that corresponds to the majority mappings in the given class (Sec. 6.).

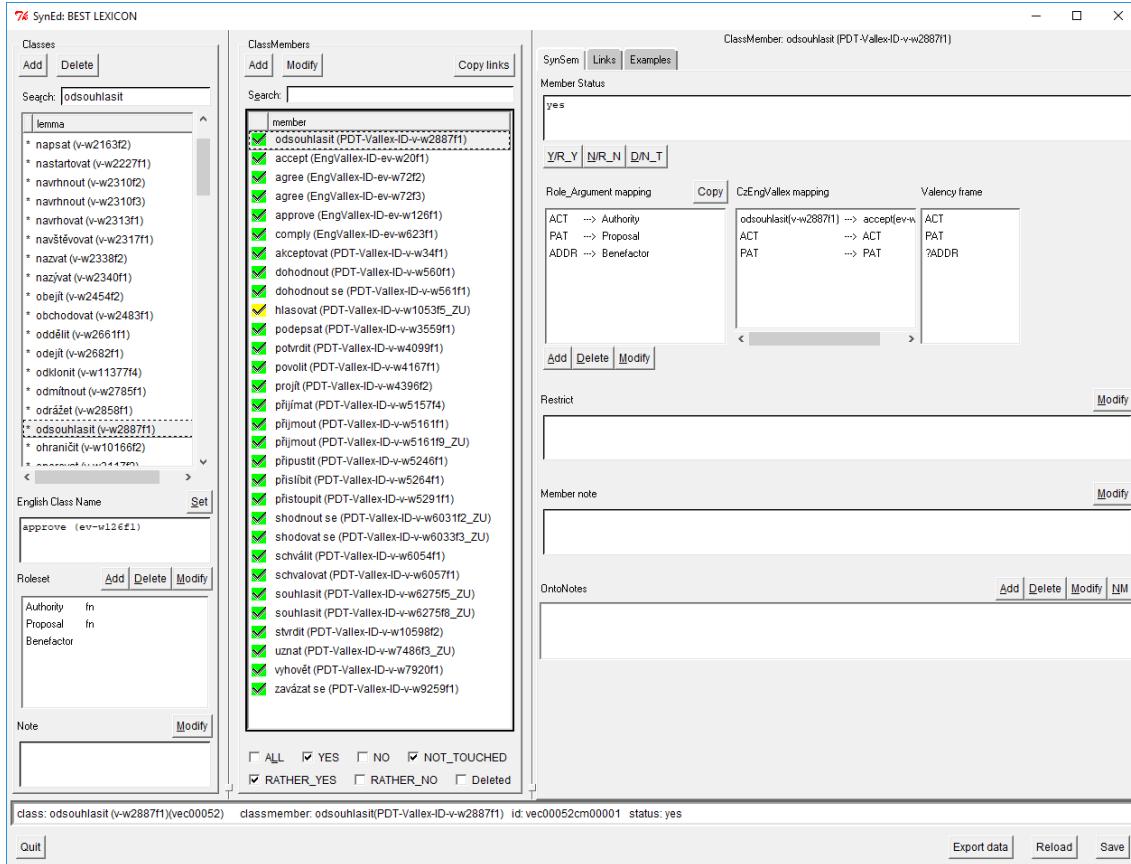


Fig. 35 SynEd – Assignment of an English verb to a class name - *English Class Name* window

9. Archiving and Distributing Annotated Data

To manage (distribute and archive) your annotated data, a special **Subversion (SVN) repository** has been set up for each annotator, called the **ÚFAL-repository**, at:

<https://svn.ms.mff.cuni.cz/svn/CzEngClass/data/anotation/<your initials>>

The instructions for installing the **TortoiseSVN client** (“**svn**”) are part of the “technical” manual that will be available to you.

To be able work as an annotator, you need to know the following operations:

Checkout - setting up the so-called *working copy of a repository* on your computer (SVN client initially copies the repository data from ÚFAL to your computer and starts tracking changes). You will only need this operation once at the beginning of your work.

Commit – projects changes made in your working copy of the repository onto the ÚFAL-repository (SVN client copies data from your working copy back to the ÚFAL-repository – it recognizes changes and copies only those). You will need this operation periodically to submit annotated data.

Update - reflects changes from the ÚFAL- repository to your working copy (SVN client copies new data and changes from the ÚFAL- repository to your working copy so that the contents of the working copy of the repository and the ÚFAL- repository match). You will need this operation periodically to get a newly allocated batch of sharp data from the UFAL-repository.

Revision - the number of the current "revision" (version) of the content of the ÚFAL-repository, each Commit will increase it (a complete history of changes in the repository is registered, we can return to the older version at any time).

Merge - SVN is designed for multi-people collaboration on the same files (multiple working copies exist). So there may occasionally be a conflict and the need to merge changes from the UFAL- repository with changes in your working copy of the repository. But for the purposes of our project, each annotator is assigned a sub-directory in the UFAL-repository that other annotators cannot access. So we will not routinely resolve conflicts.

PART II – Czech version

1. Cíl anotace

Anotace bilingválního slovníku synonym probíhá v rámci tříletého (2017 - 2019) projektu nazvaného *Contextually-based synonymy and valency of verbs in a bilingual setting*. Tento projekt je podporovaný grantovou agenturou České republiky (No. GA17-07313S).²⁷

Cílem anotace je vytvořit slovník bilingvních česko-anglických synonym. Synonyma jsou získávána z překladových textů Pražského česko-anglického závislostního korpusu. Anglické a české sloveso musí být v takovém významovém vztahu, který je z hlediska překladu funkčně adekvátní, tj. anglické a české sloveso jsou v daném(ych) kontextu(ech) synonymní a překladové sloveso adekvátně vyjadřuje funkční záměr originálu. Směřujeme k tomu, aby každá synonymní třída byla charakterizována jak významově (semantic roles), tak strukturně (valency arguments) prostřednictvím propojení (mapování) sémantických rolí a valenčních členů (Role_Argument mapping).

Zatímco valenční rámce jsou pro daná slovesa v textu již určená a zachycená v asociovaných valenčních slovnících - PDT-Vallex (pro češtinu), EngVallex (pro angličtinu) a CzEngVallex (pro překladový pár sloves), sémantické role a jejich propojení s argumenty zpracovány nejsou.

2. Úkoly anotace

Pro splnění cíle anotace je třeba splnit několik úkolů. Rozdělujeme je do dvou fází.

- V první fázi je úkolem anotace vytvořit **předběžné třídy** slovesných synonym na základě překladu z angličtiny do češtiny.
- Ve druhé (hlavní) fázi je úkolem anotace
 - vytvořit pro jednotlivé třídy „**Roleset**“ (sadu sémantických rolí) a
 - propojení (mapování) sémantických rolí a valenčních členů (**Role_Argument mapping**)
 - Jako pomůcku používáme odkazy (**linky**) do jiných lexikálních zdrojů. I tyto odkazy jsou předmětem anotace.

3. Úkoly anotace v první fázi

V první fázi potřebujeme **vytvořit předběžné třídy slovesných synonym**. Vycházíme přitom z překladových textů z angličtiny do češtiny obsažených v Pražském česko-anglickém závislostním korpusu (PCEDT²⁸).

3.1. Jak vytvořit předběžnou třídu slovesných synonym

V této fázi je úkolem anotátora ručně projít automaticky vygenerované slovesné anglické překladové protějšky k českému slovesu, a vytvořit tak kandidáty anglických slovesných členů

²⁷ <https://ufal.mff.cuni.cz/czengclass>

²⁸ <https://ufal.mff.cuni.cz/pcedt2.0/en/index.html>

jedné synonymní třídy. Při této anotaci používáme excelové soubory a řídíme se následujícím kritériem.

Anglické a české sloveso musí být ve vztahu funkčně adekvátního překladu. To znamená, že překladové sloveso adekvátně vyjadřuje funkční záměr slovesa originálu. Kontext překladu je určen na základě reálných textů PCEDT.

Příklad funkčního překladu slovesa *learn* slovesem *dozvědět se* v kontextu:

pcedt Subsequently , I_{ACT} have learned that_{PAT} a private group , ofwhich Du Pont is a part, is_{PAT} funding_{PAT} a modest program to continuedata gathering at the Scotto report stations as well as to develop moresophisticated UVB measuring instruments.

pcedt Následně jsem se dozvěděl, že_{PAT} nezávislá skupina , jejíž součástí je Du Pont , financuje_{PAT} malý program na pokračující sběr datna stanicích uvedených ve Scottově zprávě a na vývoj výkonnějších přístrojůna měření UVB záření .

Zároveň je třeba brát v úvahu, že jedno sloveso může mít více významů, které jsou v kontextu PCEDT určeny valenčním rámcem. Valenční rámce vycházejí z teorie valence ve Funkčně generativním popisu jazyka a jsou zachyceny v příslušných valenčních slovnících: CzEngVallex²⁹, PDT-Vallex³⁰ a EngVallex³¹.

Příklad valenčního rámce pro jeden slovesný význam slovesa *hradit* – v PCEDT je rámcem označen identifikátorem: *hradit* (v-w1116f1).

hradit¹ 48x,20x ACT(1) PAT(4) ?ADDR(3) ?EFF(zá+4)

(platit) za byt jim hradil všechny poplatky; zaměstnavatel h. zaměstnancům pojštění; z Fondu.MEANS nár.majetku; za vzorné chování.CAUS jim h. výdaje na cestu; za děti.SUBS h. všechny výdaje rodiče

Příklad valenčního rámce pro jeden slovesný význam slovesa *nabídnout* – v PCEDT je rámcem označen identifikátorem: *nabídnout* (v-w1940f1)

nabídnout^{109x,281x} ACT(1) PAT(4;.f;↓že;↓aby;↓at) ADDR(3) ?EFF(zá+4)

nabídl mu peníze za dům; škola n. studium absolventům středních škol; n. za auto.EFF pakatel; n. za vykonanou službu.CAUS odměnu; n. jí peníze jako odškodné.COMPL; za 100 Kč.EXT jim n. nové zboží; za poplatek.MEANS jim n. své služby; za dolary.MEANS jim n. pobyt u moře

²⁹ <https://lindat.mff.cuni.cz/services/CzEngVallex/>

³⁰ <http://lindat.mff.cuni.cz/services/PDT-Vallex/>

³¹ <http://lindat.mff.cuni.cz/services/EngVallex/>

Soubor v Excelu (Obr. 0) obsahuje dvojice class a class member. Tyto dvojice byly vybrány automaticky z překladů v PCEDT. Soubor má 8 sloupců.

Ve sloupci A (Class) je prozatímní název třídy reprezentovaný českým slovesným významem a identifikací valenčního rámce, přičemž platí, že každý slovesný význam má svůj identifikátor, který tedy rozlišuje významy. Z technických důvodů to je složitější zápis a ne pouhý index 1, 2, 3..., obvyklý ve slovnících. Např. čekat PDT-Vallex-ID-v-w31If1, jiný význam čekat: ID-v-w31If9_ZU

Ve sloupci B (Classmember) jsou potenciální členové uvedené třídy (anglická slovesa v jednom slovesném významu).

Např. pro třídu čekat (v-w31If9_ZU) je uveden jako jeden potenciální člen význam wait EngVallex-ID-ev-w3576f3 a význam wait EngVallex-ID-ev-w3576f4 jako druhý potenciální člen

Ve sloupci D (Groupings) a ve **sloupci F (CzEngVallex)** jsou odkazy do dalších zdrojů, které nám pomohou rozhodnout, zda daný člen (slovesný význam) do příslušné třídy zařadit nebo ne.

Sloupce C a E jsou ty, kam se zapisují anotace.

Do **sloupce C (Status)** zapíšeme, zda se daný člen do třídy zařadí nebo ne. Vybíráme jednu z následujících hodnot:

Y (Yes) - člen do třídy patří,
N (No) - člen do třídy nepatří.

Pokud si nejsme jistí, máme možnost vybrat také jednu z následujících hodnot. Tuto hodnotu se ale snažíme používat co nejméně.

R_Y (Rather Yes) - člen do třídy spíše patří,
R_N (Rather No) - člen do třídy spíše nepatří.

Můžeme se setkat s případem, kde byl člen do dané třídy zařazen automatickým výběrem omylem. Pokud si myslíme, že tomu tak je, přiřadíme hodnotu

D (Delete) - člen do třídy nepatří, jde o zjevnou chybu v párování sloves už v datech, odkud byly dvojice automaticky předvybrány.

Do **sloupce E (Sense)** vyplňujeme číslo anglického slovesného významu, který odpovídá významu pro danou třídu. Soubor významů pro dané anglické sloveso najdeme po rozkliknutí odkazu ve sloupci D – záznam z OntoNotes (*Groupings*). Vyplňujeme pouze v případě, že člen zařadíme do dané třídy. Člen může mít v OntoNotes jeden nebo i více významů. Pokud nenalezneme v OntoNotes odpovídající význam, zapíšeme do sloupce E hodnotu “NO”.

Poznámka: slovesa typu build_up, look_for, apod. jsou v OntoNotes zařazena ke slovesům bez podtržítka- tj. build, look. Může se stát, že odkaz v OntoNotes neotevře správnou stránku s

jednotlivými významy, pak vyhledáme správný význam na této stránce: http://verbs.colorado.edu/html_groupings/.

Sloupec **F (CzEngVallex)** obsahuje odkazy do valenčního slovníku CzEngVallex pro dvojici české sloveso (určuje *Class*) a anglické sloveso (určuje *Classmember*). Odkaz umožňuje podívat se na příklady (věty), v nichž byla obě slovesa použita. Klikem v Excelu jsme přesměrování na web, kde klikneme na *Show corpus examples* a načtou se příkladové věty.

Sloupec **G (Restriction)** je určen pro vložení omezení, které dané sloveso musí splňovat, je-li členem dané synonymní třídy. Zapisujeme jen volný text, vysvětlující v jakém kontextu lze považovat anglické sloveso za synonymní s českým slovesem. Nejsou zde žádné požadavky na formát. Důležité ale je nepoužívat odřádkování.

Sloupec **H (Notes)** je určen pro poznámky. Z důvodu automatického zpracování výsledků anotací je důležité ani zde nepoužívat víceřádkové poznámky.

A	B	C	D	E	F	G	H	
1	Class	Classmember	Status	Groupings	Sense	CzEngVallex	Restrictions	Notes
3	dát (v-w369f21)							
5		dát PDT-Vallex-ID-v-w369f21						
7		put EngVallex-ID-ev-w2449f7_u_nobody		groupings (put)		czengvallex: put(ev-w2449f7_u_nobody) - dát(v-w369f21)		
9		put EngVallex-ID-ev-w2449f15_u_nobody		groupings (put)		czengvallex: put(ev-w2449f15_u_nobody) - dát(v-w369f21)		
11		put_up EngVallex-ID-ev-w2454f1		groupings (put_up)		czengvallex: put_up(ev-w2454f1) - dát(v-w369f21)		
13		set EngVallex-ID-ev-w2900f8		groupings (set)		czengvallex: set(ev-w2900f8) - dát(v-w369f21)		
16	datovat se (v-w11235hsa_989)							
18		datovat se PDT-Vallex-ID-v-w11235hsa_989						
20		date EngVallex-ID-ev-w808f2		groupings (date)		czengvallex: date(ev-w808f2) - datovat se(v-w11235hsa_989)		
23	dit se (v-w492f2)							
25		dit se PDT-Vallex-ID-v-w492f2						
27		be EngVallex-ID-ev-w218f3		groupings (be)		czengvallex: be(ev-w218f3) - dit se(v-w492f2)		
29		happen EngVallex-ID-ev-w1548f5		groupings (happen)		czengvallex: happen(ev-w1548f5) - dit se(v-w492f2)		
32	dolahovat (v-w614f1)							

Obr. 0 Anotace příslušnosti do třídy (*Status*) a významu (*Sense*), s omezeními (*Restrictions*) a poznámkami (*Notes*)

4. Postup při výběru členů třídy

Při výběru potencionálních členů třídy volíme následující postup, rozdělený do jednotlivých kroků:

Step 0

automatický výběr 200 českých sloves (rámců) – které jsou prozatímním názvem tříd

Step 1

automatický výběr anglických slovesných překladových protějšků ke slovesům z kroku 0 určení příslušnosti do třídy (anotace statusu), celková anotace jednotlivých členů třídy, tj. Roleset, mapování, odkazy, příklady, omezení, poznámky

Step 2

automatický výběr českých protějšků ke slovesům z kroku 1 (kromě sloves z bodu 0) určení příslušnosti do třídy (anotace statusu) celková anotace jednotlivých členů třídy, tj. Roleset, mapování, odkazy, příklady, omezení, poznámky

Step 3

automatický výběr anglických slovesných překladových protějšků ke slovesům z kroku 2

Pokud budou přibývat slovesné členy, budou následovat další kroky (4, 5...).

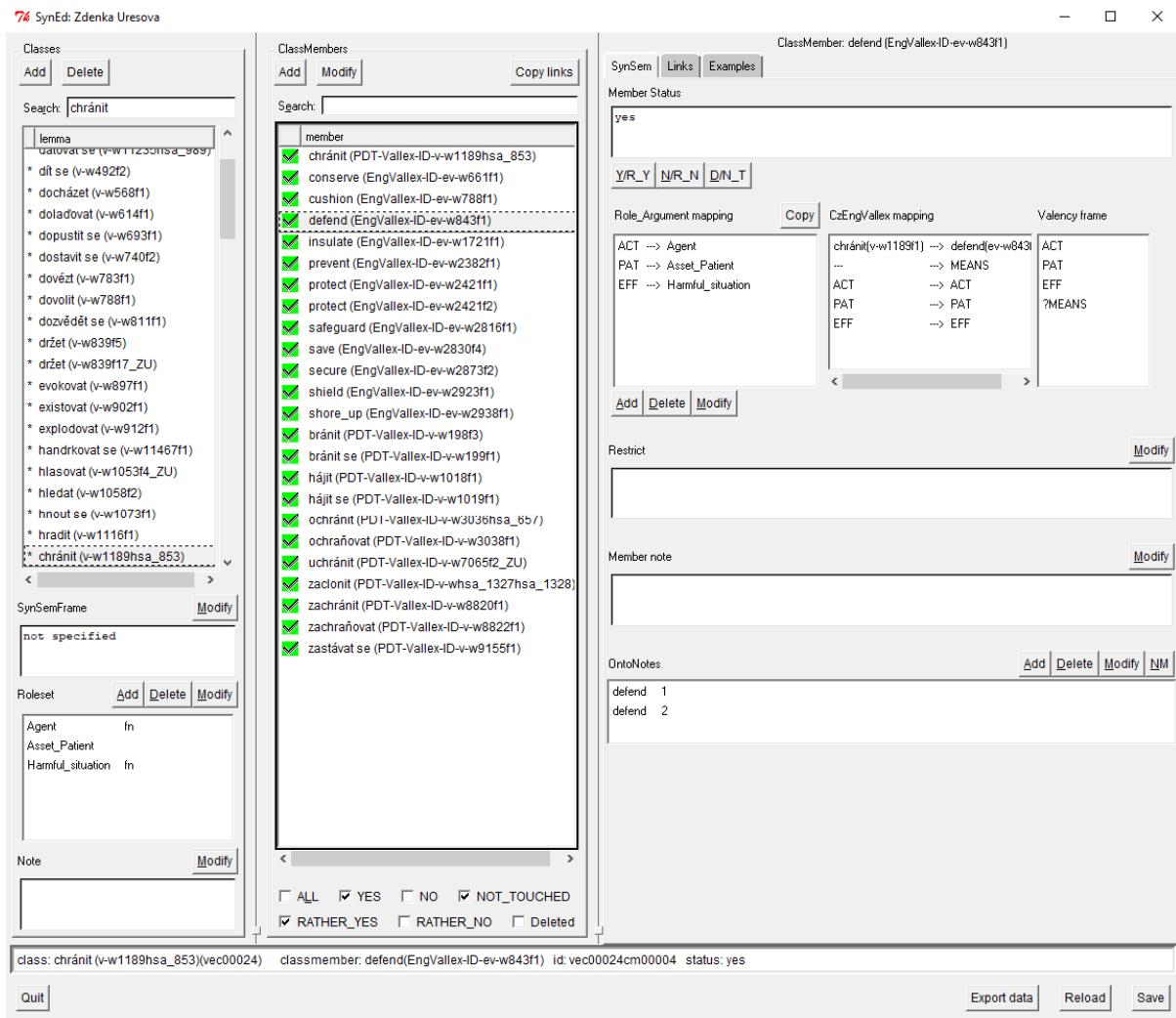
5. Úkoly anotace ve druhé fázi

Ve druhé fázi je úkolem anotace vytvořit pro jednotlivé třídy „Roleset“ (kap. 6) a **namapovat jeho členy na členy valenčního rámce jednotlivých členů třídy** (kap. 7). Jako pomůcku anotace Rolesetu a mapování používáme **odkazy do jiných lexikálních zdrojů** (kap. 5.1). I tyto odkazy jsou předmětem anotace.

Předpřipravená slovesa v jednotlivých třídách musí splňovat následující podmínky:

1. Všechna slovesa dané třídy musí mít stejnou sadu (*Roleset*) sémantických rolí.
Příklad: *Recipient, Phenomenon, Source* – třída dozvědět se (*learn, know, hear, doslychat se, naučit se, zaslechnout....*)
2. Anglický slovesný překlad by měl mít obdobnou valenci jako české sloveso v daném významu.
Příklad: *dovědět se: ACT, PAT, ORIG = learn: ACT PAT ORIG*

V této fázi anotace pracujeme ve speciálním editoru - **CzEngClass editor** (kap. 8.1.), zvaný **Synonyms Editor (SynEd)**. SynEd zachycuje (Obr. 1a, odleva doprava) *Classes*, *ClassMembers* a *syntakticko-sémantické informace (SynSem)*, *odkazy (Links)* a *příklady (Examples)*. Postup práce s editorem popisujeme průběžně v dalším textu a v kapitole 8.



Obr. 1a CzEngClass editor (*SynEd*) – zobrazení pro anotátora Zdenka Uresova, třídu *chránit* (*v-w1189hsa_853*), jejího člena *defend* (*EngVallex-ID-ev-w843f1*)

5.1. Jak vytvořit odkazy do jiných lexikálních zdrojů

Chceme, aby odkazy do jiných lexikálních zdrojů byly součástí slovníku CzEngClass, a to ze dvou důvodů. Předpokládáme, že propojení našeho slovníku s dalšími relevantními lexikálními zdroji umožní jeho interoperabilitu, a slovník tak bude mít vyšší hodnotu pro využití v počítačové lingvistice.

Zároveň díky odkazům do jiných sémantických databází získáme lepší pohled na jednotlivé

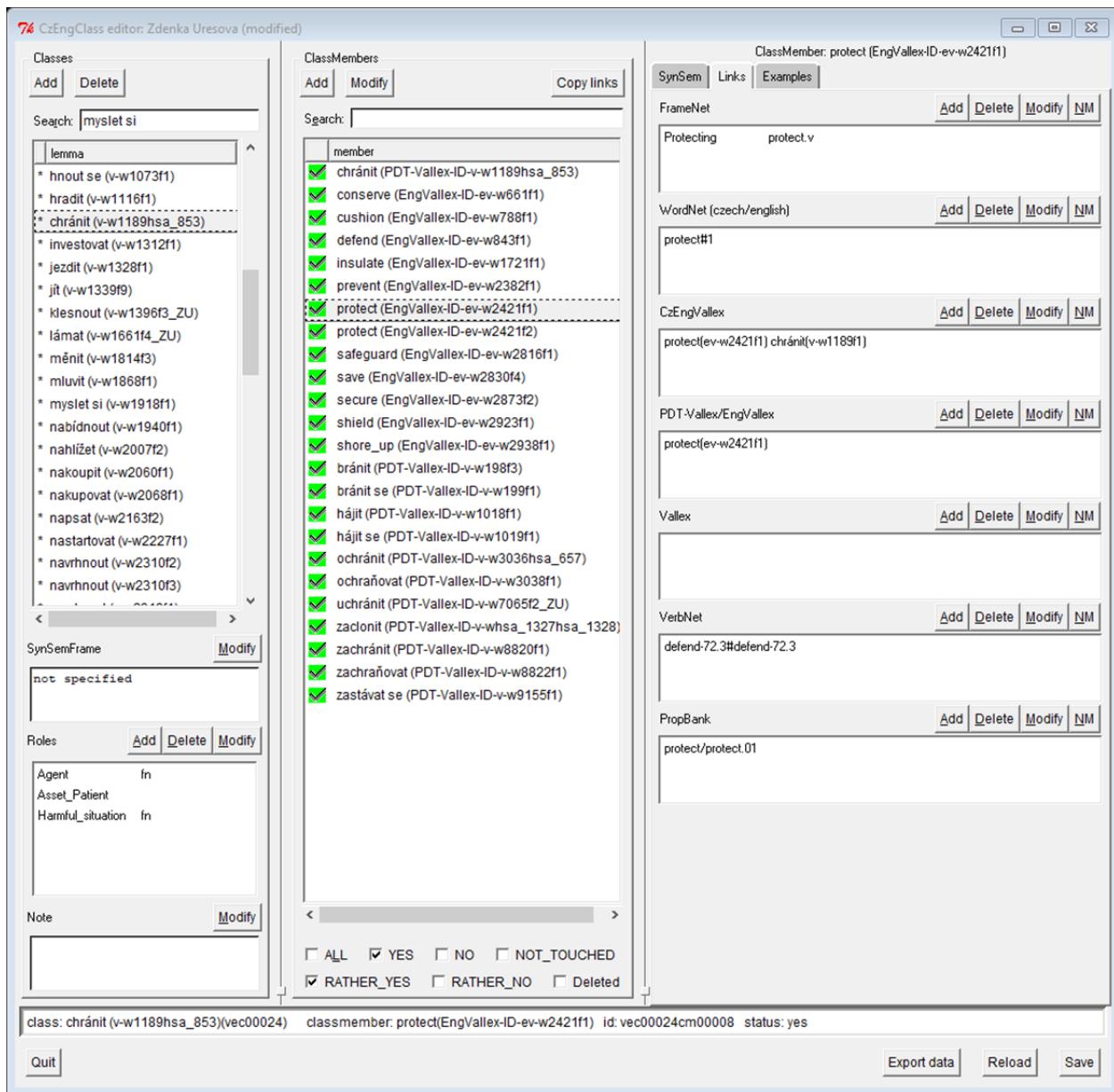
významy sloves a jejich charakteristiky. Kromě kontroly³² předpřipravených odkazů do PDT-Vallexu, EngVallexu a CzEngVallexu je úkolem anotace přiřadit v SynEdu, pod záložkou Links, k jednotlivým členům třídy (ClassMembers – CMs) následující odkazy (Obr. 1b - Links³³):

- pro české CMs: do Vallexu,
- pro anglické CMs: do OntoNotes, FrameNetu, WordNetu, VerbNetu a PropBanku.

Odkazů do jednotlivých zdrojů může být více, ale i jen jeden nebo žádný.

³² Odkazy mohou být neplatné jednak vlivem změn ve slovnících (PDT-Vallex, EngVallex, CzEngVallex) a jednak vlivem průběžné anotace slovníku CzEngClass.

³³ Vzhledem k průběžnému zachycování anotačních pravidel jsou některé obrázky již neaktuální. Necháváme je však v původní podobě, protože dokreslují postup a vývoj anotací. Na neaktuální místa v manuálu vždy upozorňujeme poznámkou „neaktuální – zkratka NeA“.



Obr. 1b. Zobrazení odkazů v editoru *SynEd* u slovesa *protect* ve třídě *chránit* (NeA)

5.1.1. Odkazy do OntoNotes

Pro anglické sloveso je klíčový odkaz do OntoNotes/Groupings. V *SynEdu* vyvoláme internetovou stránku http://verbs.colorado.edu/html_groupings/defend-v.html s odkazem do OntoNotes dvojitým klikem na příslušný anglický CM. Na základě kontextu, tj. příkladů v PCEDT (zatím – NeA – v *SynEdu* zpřístupněné v TrEdu přes záložku *Examples > Show in TrEd*), se určí správný *Sense* (Obr. 2).

Sense Number 2: protect or speak up against criticism or opposition

Commentary: Syntax Is: NP1[agent[human]] DEFEND NP2[topic[abstract]]

NOTE: generally implies a verbal endorsement of a process, state or concept

NOTE: includes legal courtroom defense.

Examples:

John defended his thesis brilliantly before the committee.

The administration is defending its economic policies.

He hasn't agreed to defend our position at tomorrow's meeting.

I won't defend her actions in this matter.

Bob defended his little sister against their mother's criticisms.

They are fighting to defend their rights.

He vigorously defended his point of view.

That law firm has a reputation for defending members of the mob.

The famous attorney will defend her in the high-profile murder case.

Mappings:

VerbNet: defend-72.2

FrameNet: Justifying

PropBank: defend.01

WordNet 3.0 Sense Numbers: 1, 5, 6, 7

Obr. 2 Vybraný sense z OntoNotes pro sloveso *defend*

U každého významu v OntoNotes najdeme pod *Mappings* odkazy daného významu slovesa do VerbNetu, FrameNetu, PropBanku a WordNetu (např. pro sloveso *defend*, *sense 2*):

Mappings:

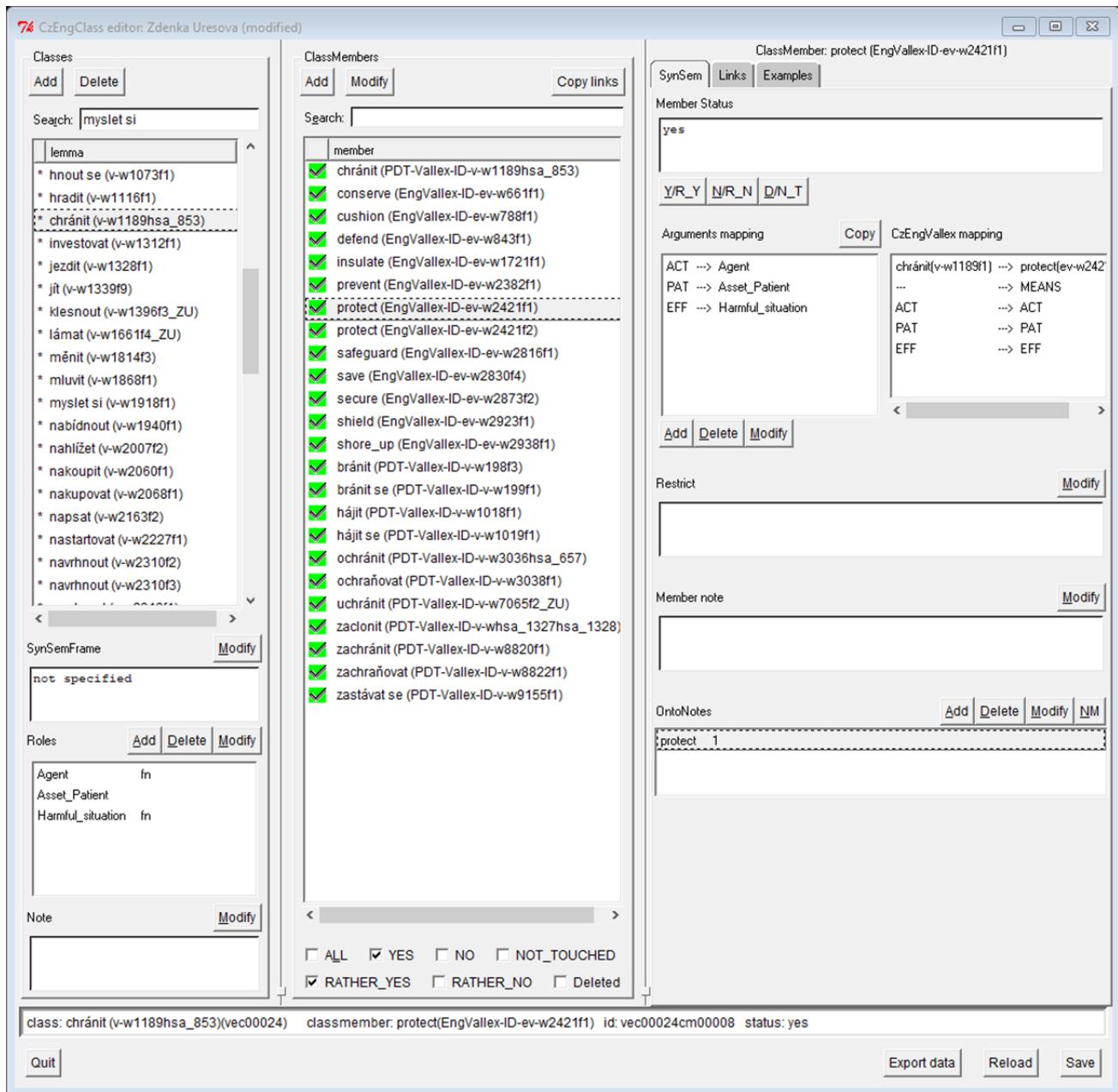
VerbNet: defend-72.2

FrameNet: Justifying

PropBank: defend.01

WordNet 3.0 Sense Numbers: 1, 5, 6, 7

Zde uvedené odkazy použijeme jako výchozí pomůcku pro anotaci odkazů v CzEngClass slovníku. Pozor, je třeba všechny odkazy zkontolovat; nemusí totiž platit (změnila se verze anotací v propojených zdrojích, nebo je anotace chybná z jiného důvodu). Zobrazení odkazů do OntoNotes v SynEd zachycuje Obr. 3 (sloveso *protect* – sense 1, vpravo dole). Odkazů může být 0 – zadáme *No Mappings* (NM), 1, nebo více. Pro přidávání odkazů použijeme tlačítko *Add*. Stejně jako jiné položky v editoru lze i tuto modifikovat (*Modify*), nebo vymazat (*Delete*).



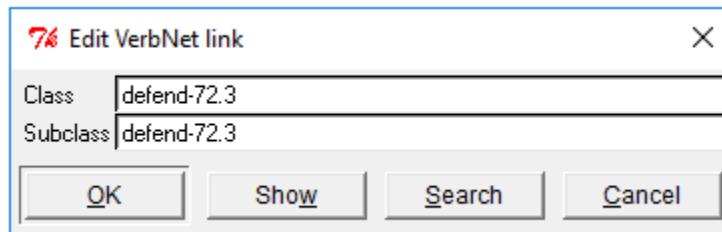
Obr. 3. Zobrazení odkazů pro CM *protect* do Ontonotes v *SynEdu* ((NeA)

Mappings v OntoNotes mohou také chybět.

5.1.2. Odkazy do VerbNetu

Během anotování je výhodné dále pokračovat přiřazováním odkazů do VerbNetu. Pokud pro daný CM známe odkaz do VerbNetu z *Mappings* uvedených u OntoNotes, tlačítkem *Add* odkaz

přířadíme a přes tlačítko *Show* zkontrolujeme, zda je správně (Obr. 4). Tlačítko *Show* nás zavede do anotace daného významu slovesa ve VerbNetu³⁴.



Obr. 4 Editace odkazů do VerbNetu pro sloveso *defend*

Pokud *Show* nefunguje, použijeme tlačítko *Search* (Obr. 4), které nás zavede na stránku Semlinku.³⁵ Zde přes tlačítko *Search* (Obr. 5) najdeme dané sloveso (*defend*).³⁶

Obr. 5. Anotace pro sloveso *defend* v SemLinku

³⁴ <http://verbs.colorado.edu/verb-index/vn3.3/vn/defend-72.3.php#defend-72.3>

³⁵ <http://verbs.colorado.edu/verb-index/vn3.3/>

³⁶ <http://verbs.colorado.edu/verb-index/vn3.3/search>

Rozkliknutím třídy *defend-72.3* získáme další detaily anotace: **DEFEND (wn 1, 6, 7; g 1, 2, 3)**, které můžeme využít. Zkratky v závorce za slovesem jsou odkazy do FrameNetu (zde chybí), WordNetu (**wn 1, 6, 7**) a tzv. Groupings (**g 1, 2, 3**). I zde je možné rozkliknutím zobrazit příslušný odkaz.

Pokud pro daný CM odkaz z OntoNotes neexistuje, nebo nedokážeme žádný z uvedených vybrat (v *SynEdu* je pro oba případy uvedeno *NM*), tlačítkem *Add* (Obr. 6),



Obr. 6 Okno pro přiřazování odkazů do VerbNetu – editační tlačítka vpravo nahoře

otevřeme Unified Verb Index³⁷, kde pomocí *Search* vyhledáme dané sloveso. Zobrazení pro sloveso *defend* na Obr. 5, pro sloveso *save* na Obr. 7.

Obr. 7 Vyhledané sloveso *save* v SemLinku

Na levé straně obrazovky pod *VerbNet Members*, případně *VerbNet Classes* najdeme zobrazené anotace slovesa v SemLinku (např. *save* zařazeno do třídy *bill.54.5* a *get-13.5.1-1*).

³⁷ <http://verbs.colorado.edu/verb-index/vn3.3/>

Pokud žádná třída z VerbNetu nevyhovuje našemu CM, zaneseme k příslušnému slovesu (jako v případě slovesa *save*, aby CM třídy *chránit*, Obr. 7) status odkazu do Verbnetu *No Mapping* - přes tlačítko *NM* (Obr. 6).

Pokud můžeme hledané sloveso zařadit do některé z červeně zobrazených tříd VerbNetu, vybranou třídu (*VerbNet Members*, případně *VerbNet Classes*) rozklikneme (Obr. 8 pro sloveso *defend*) a její název vložíme jako odkaz přes tlačítko *Add* (Obr. 6) do *SynEdu* (název třídy lze pro usnadnění kopírovat). Dále můžeme zkontolovat, zda ve VerbNetu uvedené odkazy (pro sloveso *defend*: **DEFEND (WN 1, 6, 7; G 1, 2, 3)**; **G 1, 2, 3**) vyhovují naší anotaci odkazů. Pokud ano, můžeme odkazy rozklikat a využít ke vkládání dalších odkazů do *SynEdu*.

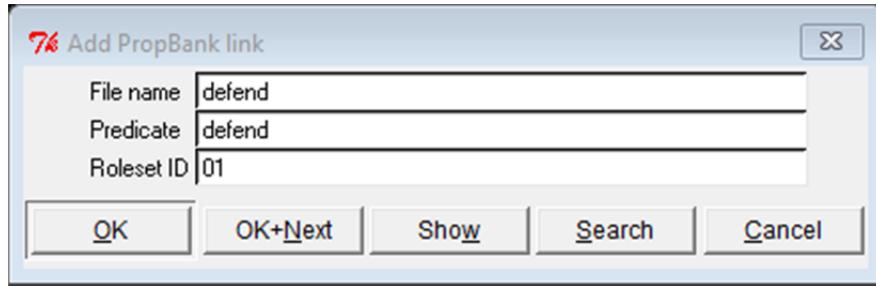
Obr. 8 Anotace slovesa *defend* ve třídě defend-72.3

5.1.3 Odkazy do PropBanku

Pro přiřazení odkazu/ů do PropBanku lze využít nápovědu z *Mappings* uvedených u OntoNotes (Obr. 2 *defend*): **PropBank: defend.01**. Odkaz **defend.01** přiřadíme obdobným způsobem jako odkaz do VerbNetu, použijeme ale okno *PropBank* (Obr. 9) a zapisujeme do *Roleset ID* okna přesně, tedy např. 01, nikoli 1 (Obr. 10).



Obr. 9 Okno pro přiřazování odkazů do PropBanku – editační tlačítka vpravo nahoře



Obr. 10 Editace odkazů do PropBanku pro sloveso *defend*

Pro kontrolu přes tlačítko *Show* otevřeme stránky anotace příslušného slovesa v PropBanku.³⁸

Můžeme využít také údaj ze SemLinku (Obr. 5). Zde lze najít anotaci příslušného slovesa v pravé části pod hlavičkou PropBank.

Pokud žádnou nápovědu nemáme, otevřeme opět přes tlačítko *Search* (Obr. 10) stránky SemLinku³⁹, kde vyhledáme příslušné sloveso a jeho anotaci v PropBanku. Sloveso je zde podchyceno i v případě, že není anotováno ve VerbNetu (Obr. 11 – pro sloveso *cushion*).

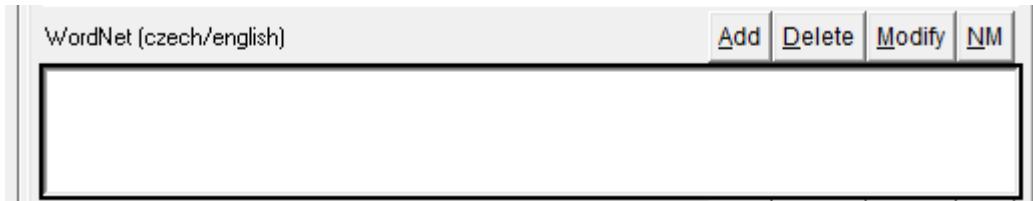
Obr. 11 Odkaz pro sloveso *cushion* do PropBanku přes databázi SemLink

5.1.4. Odkazy do WordNetu

³⁸ <http://verbs.colorado.edu/propbank/framesets-english-aliases/defend.html>

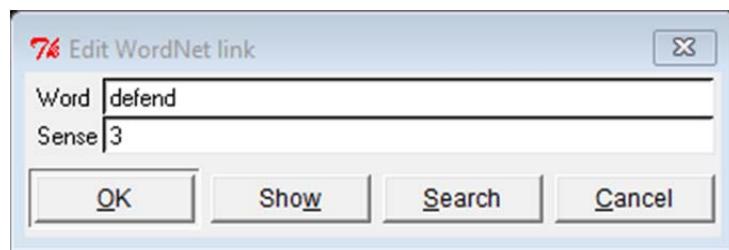
³⁹ <http://verbs.colorado.edu/verb-index/vn3.3/>

Pro přiřazení odkazu/ů do anglického WordNetu lze opět využít nápovědu z *Mappings* uvedených u OntoNotes (Obr. 2 *defend*): WordNet 3.0 Sense Numbers: 1, 5, 6, 7. Odkaz doplňujeme do okna *WordNet* (Obr. 12),



Obr. 12 Okno pro přiřazování odkazů do WordNetu – editační tlačítka vpravo nahoře

kde do *Sense* okna (Obr. 13) zapisujeme číslo významu daného slovesa, odpovídající významu uvedenému v databázi WordNet⁴⁰ (Obr. 14). Zapisujeme jen číslo (3), nikoli číslo s křížkem (#3).



Obr. 13 Editace odkazů do WordNetu pro sloveso *defend*

⁴⁰ <http://wordnetweb.princeton.edu/perl/webwn?&o7=1&s=defend>

WordNet Search - 3.1

- [WordNet home page](#) - [Glossary](#) - [Help](#)

Word to search for:

Display Options: ▾

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

Display options for sense: (gloss) "an example sentence"

Display options for word: word#sense number

Verb

- S: (v) **defend#1**, [support#8](#), [fend for#1](#) (argue or speak in defense of) "She supported the motion to strike"
- S: (v) **defend#2** (be on the defensive; act against an attack)
- S: (v) **defend#3**, [guard#3](#), [hold#20](#) (protect against a challenge or attack) "Hold that position behind the trees!"; "Hold the bridge against the enemy's attacks"
- S: (v) [fight#2](#), [oppose#2](#), [fight back#2](#), [fight down#1](#), **defend#4** (fight against or resist strongly) "The senator said he would oppose the bill"; "Don't fight it!"
- S: (v) [champion#1](#), **defend#5** (protect or fight for as a champion)
- S: (v) **defend#6**, [represent#8](#) (be the defense counsel for someone in a trial) "Ms. Smith will represent the defendant"
- S: (v) [maintain#8](#), **defend#7** (state or assert) "He maintained his innocence"

Obr. 14 Databáze WordNet s vyhledanými významy slovesa *defend*

Pokud žádnou nápovědu nemáme, hledáme odkaz pro CM rovnou v databázi WordNet, kterou opět jednoduše otevřeme tlačítkem *Search* v okně pro přiřazování odkazů do WordNetu⁴¹ (Obr. 12).

5.1.5. Odkazy do FrameNetu

Z pohledu sémantických rolí (SRs) je pro nás nejdůležitějším zdrojem a inspirací elektronická databáze FrameNet⁴², v níž se pracuje se sémantickými rolemi (frame elements - FEs) uvnitř sémantických rámců (semantic frames).

⁴¹ <http://wordnetweb.princeton.edu/perl/webwn>

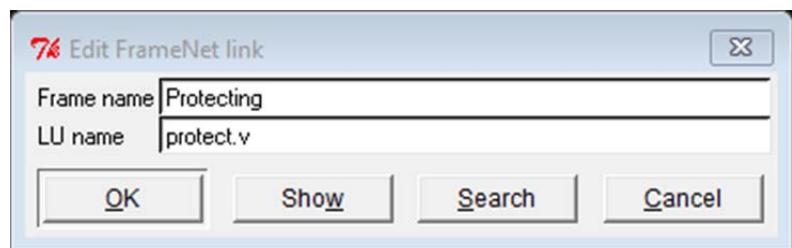
⁴² <https://framenet.icsi.berkeley.edu/fndrupal/about>

Se SRs pracuje i VerbNet⁴³, můžeme se jimi také inspirovat. Tím, že vytvoříme pro CM odkaz do sém. rámce/ů ve FrameNetu, získáme také možnost využít SRs, které jsou v tomto rámci přiřazené. Odkaz/y do FrameNetu zapisujeme do okna FrameNet (Obr. 15).



Obr. 15 Okno pro přiřazování odkazů do FrameNetu – editační tlačítka vpravo nahoře

Postupujeme obdobně jako u výše popsaných odkazů. Editační okno FrameNetu (Obr. 16) obsahuje okno pro název sémantického rámce ve FrameNetu (*Frame name*) a okno pro název lexikální jednotky (*LU name*), které se odkaz týká.



Obr. 16 Editační okno FrameNetu

Máme-li pro CM (*protect*, třída *chránit*) díky mapování uvedeném u významu anotovaného CM v OntoNotes již odkaz, jen ho přepíšeme do editačního okna FrameNetu a přes tlačítko *Show* ověříme jeho platnost v databázi FrameNet (Obr. 16). Jméno sémantického rámce (*Protecting*) je ve FrameNet databázi uvedeno v pravém horním rohu (Obr. 17).

⁴³ <https://verbs.colorado.edu/~mpalmer/projects/verbnet.html#thetaroles>

Secure | <https://framenet2.icsi.berkeley.edu/fnReports/data/lu/lu12789.xml>

Apps Zimbra Imported From Firefox Google Maps Institute of Formal and Applied Linguistics internalczengvallex synology iDNES.cz – zprávy, kteří IDOS - MHD Praha ... Other bookmarks Lexical Entry Protecting

Annotation

protect.v

Frame Element	Core Type
Asset	Core
Beneficiary	Extra-Thematic
Danger	Core
Degree	Extra-Thematic
Depictive	Extra-Thematic
Descriptor	Extra-Thematic
Duration	Extra-Thematic
Event description	Extra-Thematic
Explanation	Extra-Thematic
Instrument	Peripheral
Manner	Peripheral
Means	Peripheral
Place	Peripheral
Protection	Core
Time	Peripheral

[Turn Colors Off](#)

- Wenemy forces defense-(1)
 1. **The lumbering carts and chariots** are **PROTECTED** by outriders mounted on giant wolves who patrol the area in front of the advancing tribe , probing for enemies and scouting for small settlements to loot [CNI]
 2. If the water table is within 15 m of the surface , the cheapest method is to dig a well by hand and **PROTECT** the walls from collapse by building a masonry lining .[CNI]
 3. It was **an old graveyard** , **PROTECTED** by the high walls from the bustle and intrusion of the outside world .
- NP-T-NP-PPfrom-(1)
 1. **These regulations** are designed to **PROTECT** people against immediate or delayed health risks from substances which may have an adverse effect on their health ; for instance , disinfectants such as glutaraldehyde .
 2. **IN THE MENDIPS** **A SHEEP FARMER** IS STRUGGLING TO **PROTECT** HIS LAMBS FROM THE FOXES THAT WATCH AND PREY .
 3. **Paint** **PROTECTS** your car from rust .
 4. **Tod Richardson** **PROTECTS** himself from burglars .
 5. But **China** is also **PROTECTING** **North Korea** from the wrath of the United Nations Security Council .
- NP-T-PPfrom-(1)
 1. **Children** must be **PROTECTED** from indoctrination from such sources .[CNI]
 2. Scottish parents are apparently even more likely to do this -- **one in three parents** never **PROTECT** their children from its harmful rays despite the fact that many of them have

Obr. 17 Odkaz do FrameNetu pro sloveso *protect*, třída *chránit*

Rozkliknutím jména semantického rámce otevřeme jeho definici ve FrameNetu (Obr. 18). Vedle definice jsou zde přehledně uvedeny sémantické role (Frame Elements - FEs).

Secure | <https://framenet2.icsi.berkeley.edu/fnReports/data/frameIndex.xml?frame=Protecting&banner=1>

Frame Index Protecting Lexical Unit Index

Definition:

Some **Protection** prevents a **Danger** from harming an **Asset**.
The rubber on her sneakers PROTECTED her from the mud.

The coral reefs **SHIELDED** the coastline from the worst of the waves.

A growing number of sustainable farmers are preserving agricultural variety and **PROTECTING** biodiversity by raising 'heritage' or 'heirloom' varieties.

The man responsible for this entire ordeal **cowardly SHIELDED** himself from justice with the little boy's struggling body.

FEs:

Core:

Asset [Asset] Something desirable possessed by or directly associated with the **Protection** which might be lost or damaged.
Almost 10 years later, that agreement **INSULATED** Microsoft from liability on copyright claims over the graphical user interface (GUI).

Danger [Danger] A situation that could damage the **Asset**.
Other nations tended to muffle or modulate their occasional annoyance with the country that was their **SHIELD** against the Soviet threat.

Protection [pro] The person, entity, or action that prevents harm to an **Asset**.
The tabloid's deal with the woman **SHIELDED** Schwarzenegger.

Non-Core:

Beneficiary [i] FN: This extra-thematic FE applies to participants that derive a benefit from the protecting event.

Degree [i] This frame element selects some gradable attribute and modifies the expected value for it.

Depictive [i] FN: This FE describes a participant of the state of affairs introduced by the target as being in some state during the action.

Descriptor [des] A description or characteristic of the **Protection**.

Duration [i] FN: This FE denotes the length of time in which the **Asset** is protected.

Event description [eve] This FE is used for phrases that describe the protecting event as a whole. It applies to appraisals of the event, typically a judgment of its expectedness. The notion of appraisal requires a factive portrayal of the event. It also applies to descriptions of the protecting event as a role filler in another frame.

Obr. 18. Definice a role pro sémantický rámec *Protecting* ve FrameNetu

Pokud s odkazem souhlasíme, máme anotaci hotovu. Pokud ale odkaz pro daný CM k dispozici nemáme, nebo s ním nesouhlasíme, postupujeme následovně.

Tlačítkem *Search* v editačním okně FrameNetu (Obr. 16) otevřeme index lexikálních jednotek FrameNetu⁴⁴, kde do políčka *Search* zapíšeme hledaný CM (Obr. 19). Pokud je CM ve FrameNetu obsažen, najdeme v seznamu vyhledaných přiřazených rámců ten, který nejlépe vyhovuje naší anotaci (pro sloveso *protect* vybereme *Protecting*). Někdy lze CM odkázat na víc vyhovujících rámců; pak všechny zapíšeme do editačního okna přes tlačítko *Add*, popřípadě přes tlačítko *OK + Next*. Pokud odkaz nenajdeme, nebo žádný nevyhovuje, použijeme status *NM*.

⁴⁴ <https://framenet.icsi.berkeley.edu/fndrupal/luIndex>

The screenshot shows a web browser window with the URL <https://framenet.icsi.berkeley.edu/fndrupal/lulIndex>. The page title is "FrameNet Index of Lexical Units". A search bar at the top contains the word "protect". Below the search bar, there is a list of search results:

- protect.v ([Protecting](#)) [Finished_Initial Lexical Entry Annotation](#)
- protection [entity]n ([Protecting](#)) [Created Lexical Entry Annotation](#)
- protection [event]n ([Protecting](#)) [Created Lexical Entry Annotation](#)

Obr. 19 Vyhledávání ve FrameNetu

6. Jak vytvořit Roleset

Roleset zachycuje význam synonymní bilingvní třídy; je **definován sadou sémantických rolí** (Semantic roles - SRs), která je společná všem členům (CMs) jedné synonymní třídy. Tuto sadu rolí vytvoříme na základě propojení (namapování) sémantických rolí s valenčními členy rámce jednotlivých CMs synonymní třídy (kap. 7).

Příklady Rolesetu u třídy:

,,klesnout-decline": *Item, Initial_value, Final_value, Difference*

,,nabídnout-offer": *Offerer, Recipient, Entity_offered, Entity_received*

,,chránit-defend": *Agent, Asset_Patient, Harmful_situation*

Anotace Rolesetu velice úzce souvisí s namapováním valenčních členů na jednotlivé role (kap. 6), protože možnost namapovat valenční chování slovesa v dané třídě na role je zároveň určuje. Rozhodnutí, které role pro danou synonymní třídu vybrat není jednoduchá záležitost, která by byla proveditelná hned po anotaci jednoho člena třídy. Často si až v průběhu anotace všech potenciálních členů třídy díky odkazům ujasníme sémantický charakter jednotlivých členů třídy a také si ujasníme, které SRs přiřadíme argumentům z valenčních rámců jednotlivých CMs.

Každá synonymní třída má své **Typické Mapování Rolesetu a Argumentů** valenčního rámce (**TypRAM**), které odpovídá většinovému mapování v dané třídě. Zároveň **typické mapování**

odpovídá mapování uvedenému jak u českého, tak u anglického slovesa vybraného do názvu třídy (o anotaci názvu třídy kap. 8.1.3.).

Příklady TypRAMu u třídy:

klesnout (*v-w1396f3_ZU*) – *decline* (*ev-w829f2*):

ACT > *Item*,
ORIG > *Initial_value*,
PAT > *Final_value*,
DIFF > *Difference*

nabídnout (*v-w1189hsa_853*) – *offer* (*ev-w2105f1*):

ACT > *Offerer*,
ADDR > *Recipient*,
PAT > *Entity_offered*,
EFF > *Entity_received*

chránit (*v-w1189hsa_853*) – *defend* (*ev-w843f1*):

ACT > *Agent*,
PAT > *Asset_Patient*,
EFF > *Harmful_situation*

6.1. Repertoár sémantických rolí v CzEngClass

Repertoár sémantických rolí v CzEngClass slovníku není zatím pevně stanoven, jeho tvorba je rovněž součástí anotace. Existují různé přístupy k inventáři rolí, který má podle jednotlivých přístupů různou velikost: od velkého (cca 1300 FEs ve FrameNetu), přes střední (36 theta rolí ve VerbNetu), až malému počtu (2 proto-role v koncepci Dowty 91, nebo 2 macrorole v koncepci Van Valina 99⁴⁵).

Předpokládáme, že repertoár rolí v CzEngClass (*CzEngClass-Roles-Inventory, CERI*), vzhledem k účelu, pro jaký je používáme (synonymní bilingvní třídy), nebude doslova kopírovat žádnou již existující koncepci vytvořenou vždy pro jeden jazyk. Nicméně se zřetelem k dlouholeté tradici práce se SRs nemáme ambici vytvářet zcela nová pojmenování rolí, ale inspirujeme se pojmenováními už zavedenými.

⁴⁵ Dowty, David R. 1991. Thematic proto-roles and argument selection. *Language* 67:547–619
Van Valin, Robert D., Jr. 1999. Generalized semantic roles and the syntax-semantics interface. In *Empirical issues in formal syntax and semantics*. Vol. 2. Edited by Francis Corblin, Carmen Dobrovie-Sorin, and Jean-Marie Marandin, 373–389. The Hague: Thesus.

Východiskem pro **CERI** je anotace SRs ve FrameNetu, a to zejména proto, že jsou vytvářeny pro konkrétní sémantické rámce, které se mohou vztahovat k významově příbuzným slovesům. Od FrameNetu se ale plánujeme významně lišit v počtu rolí zařazených do CERI. Snažíme se mít rolí co nejméně, ale zároveň chceme, aby vystihovaly význam. Představujeme si repertoár o cca 200 SRs. V současném verzi CERI je v seznamu pro výběr role (Obr. 22) k dispozici 89 SRs převzatých z FrameNetu (Tab. 1) a 23 nově vytvořených SRs (Tab. 2).

Act	Interlocutor_2
Action	Issue
Activity	Item
Addressee	Label
Agent	Manner
Agreement	Medium
Appraisal	Message
Area	Money
Arguer_1	Money_owner
Arguer_2	Offerer
Assessor	Opinion
Asset	Origin
Authority	Participant_1
Buyer	Participant_2
Cognizer	Patient
Cognizer_agent	Perceiver
Comparator	Perceiver_agentive
Complainier	Perpetrator
Complaint	Phenomenon
Components	Prize
Content	Profiled_item
Created_entity	Project
Creator	Proposal
Difference	Proposition
Driver	Protagonist
Earner	Purpose
Earnings	Recipient
Entity	Requirement
Seller	Support
Sender	Supported
Signatory	Supporter
Sought_entity	Target_currency
Source	Theme
Source_currency	Topic
Speaker	Undesirable_situation
Standard_item	Value

State	
Stimulus	

Tab. 1 Seznam SRs převzatých z FrameNetu (aktuální verze v SynEdu)

Activity Event Process	Driver Entity
Agent Entity	Entity offered
Area 1	Entity received
Area 2	Goal Destination
Asset Patient	Means Vehicle
Asset Theme	Patient Entity
Audience Addressee	Phenomenon Fact
Benefactive	Proposition Fact Phenomenon
Benefactor	Recipient Place
Cognizer Inspector	Speaker Agent
Cognizer Item	Speaker Medium
Delivered entity	Topic Message
	Transporter

Tab. 2 Seznam nově vytvořených SRs, NEpřevzatých z FrameNetu (aktuální verze v SynEdu)

Vzhledem k tomu, že v této fázi tvorby slovníku **CERI** teprve vytváříme, nepovažujeme ani pravidla pro vytváření nových rolí, ani stávající seznam za definitivní. Uvědomujeme si, že anotace většího počtu sloves může nejenom seznam, ale i pravidla anotace SRs ovlivnit.

6.2. Anotace sémantických rolí v CzEngClass

Anotace SRs ve slovníku CzEngClass znamená, že přiřazujeme SRs do *Rolesetu* (Obr. 20) jednotlivých synonymních tříd, čímž zároveň vytváříme repertoár SRs ve slovníku CzEngClass (CERI).

Při anotaci SRs do *Rolesetu* rozlišujeme dvě situace:

Situace 1: máme pro daný CM k dispozici odkaz na sémantický rámec ve FrameNetu, pak přebíráme do CERI SRs z tohoto sém. rámce.

Při editaci rolí v *SynEdu* tyto role

- bud' vybíráme ze seznamu předpřipravených rolí (Obr. 21 a Tab. 1),
- nebo je do seznamu doplníme (Obr. 22 – *Add role*, při definování role nezapomeneme zakliknout *From FrameNet* – Obr. 23a),

Situace 2: pro daný CM nemáme k dispozici odkaz na sémantický rámec ve FrameNetu, pak se bud' inspirujeme rolemi uvedenými ve FrameNetu u příbuzných sloves,

Při editaci v *SynEdu* tyto role opět

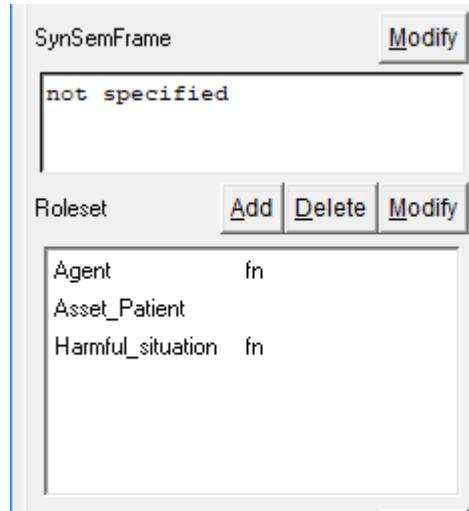
- bud' vybíráme ze seznamu předpřipravených rolí (Obr. 21 a Tab. 1),

- nebo je do seznamu doplníme (Obr. 22 – *Add role*, při definování role nezapomeneme zakliknout *From FrameNet* – Obr. 23a)).

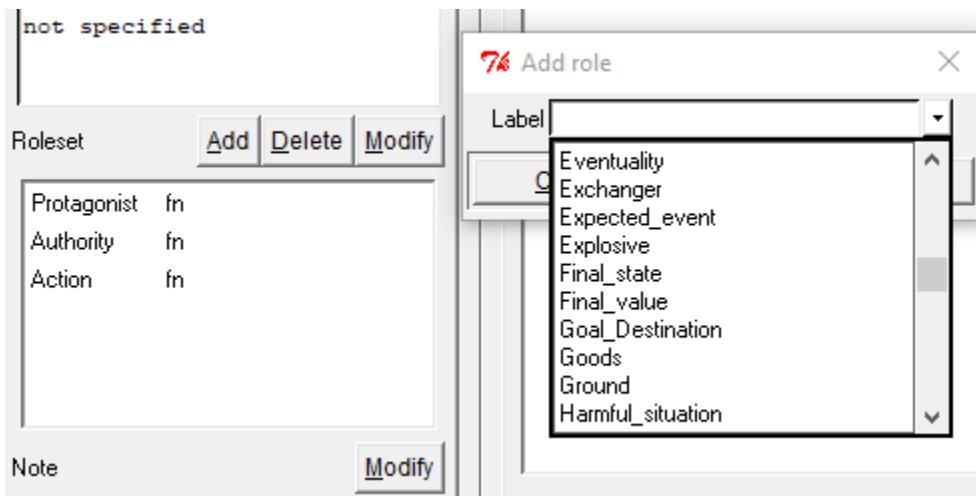
nebo tvoříme vlastní nová pojmenování SRs.

Při editaci v *SynEdu* tyto role

- bud' vybíráme ze seznamu předpřipravených rolí (Obr. 21 a Tab. 2)⁴⁶,
- nebo je do seznamu doplníme (Obr. 22 – *Add role*, při definování role nezapomeneme NEzakliknout *From FrameNet* – Obr. 23b)).

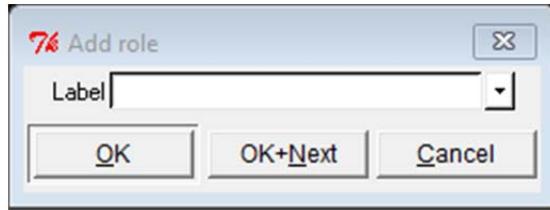


Obr. 20 Okno *Roleset* v *SynEdu*

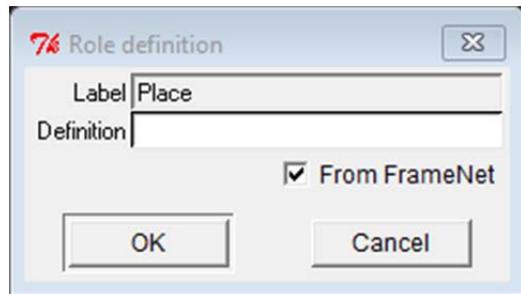


Obr. 21 Výběr SRs z CERI

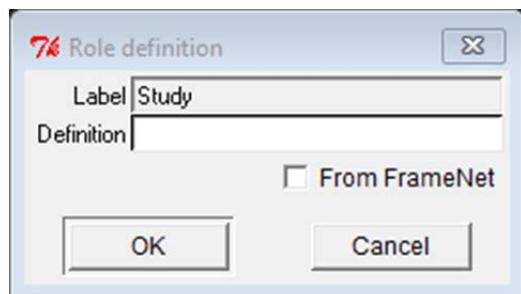
⁴⁶ V okně *Add Role* není v seznamu rolí rozlišováno, která role je z FrameNetu a která ne. Původ role je indikován pouze v okně *Roleset* (značkou *fn* pro role z FrameNetu) po pravé straně role.



Obr. 22 Přidávání nové SR



Obr. 23a Definice nové role- přejaté z FrameNetu (zakliknout *From FrameNet*)



Obr. 23b Definice nové role- nepřejaté z FrameNetu (nezakliknout *From FrameNet*)

Zatím definici sémantické role neanotujeme. Budeme dopňovat později.

Situace 1 - máme pro daný CM k dispozici odkaz na sémantický rámec ve FrameNetu

Pokud více CMs synonymní třídy X obsahuje odkaz na společný nebo podobný sémantický rámec ve FrameNetu, je anotace SRs v Rolesetu jednodušší. Zkontrolujeme repertoár FEs v těchto společných rámcích a vybereme ze seznamu FEs (Obr. 21) ty SRs (resp. FEs), které nejlépe vyhovující naší anotaci. Je třeba brát v úvahu to, že FEs, coby reprezentanti sémantické informace v daném sémantickém rámci, nejsou chápány jako univerzální sémantické role, ale jako specifické pro daný rámec. Proto například u CMs třídy *chránit* najdeme pro namapování valenčního aktoru (ACT) několik možných SRs, podle toho, do jakého sémantického rámce jednotliví CMs ve FrameNetu odkazují.

Příklady:

- Classmember *save* (*ev-w2830f4*) odkazuje do sémantického rámce *Rescuing*, kde aktoru odpovídá *Agent*:

https://framenet2.icsi.berkeley.edu/fnReports/data/frameIndex.xml?frame=Rescuing&ba_nner=

Rescuing

An **Agent** saves a **Patient** or an **Asset** from a **Harmful** situation.

Israeli commandos RESCUED **the hostages** and left.

- Classmember *defend* (*ev-w843f1*) odkazuje do sémantického rámce *Defending*, kde aktoru odpovídá *Defender*:

https://framenet2.icsi.berkeley.edu/fnReports/data/frameIndex.xml?frame=Defending&ba_nner=

Defending

A **Defender** responds to an **Assailant**'s attack on a **Victim** by directly, and usually violently, countering it.

Samantha **stalwartly** DEFENDED **the gateway** **from the wraiths**, **crouching low over Stephen's prone form**.

- Classmember *protect* (*ev-w242If1*) odkazuje do sémantického rámce *Protecting*, kde aktoru odpovídá *Protection*:

https://framenet2.icsi.berkeley.edu/fnReports/data/frameIndex.xml?frame=Protecting&ba_nner=

Protecting

Some **Protection** prevents a **Danger** from harming an **Asset**.

The rubber on her sneakers PROTECTED **her** **from the mud**.

Na základě porovnání FEs vybereme pro daný valenční člen nevhodnější „nálepku“ pro sémantickou roli. Pro uvedený příklad třídy *chránit* považujeme za nevhodnější nálepku pro sémantickou roli namapovanou na aktor sémantickou roli *Agent*.

Pokud v CERI roli, kterou chceme z FrameNetu přiřadit, nenajdeme, doplníme ji (Obr. 22 – *Add role*, při definování role⁴⁷ nezapomeneme zakliknout *From FrameNet* – Obr. 23a).

Roleset pro danou synonymní třídu můžeme editovat dvojím způsobem:

1. Budě jednotlivé SRs zaneseme v *SynEdu* rovnou do okna *Roleset* (Obr. 20) pomocí editačního tlačítka *Add*,

⁴⁷ Definici rolí zatím neanotujeme.

2. nebo SRs zanášíme nejprve při anotaci mapování SRs na valenční členy u jednotlivých CMs do okna *Role Argument mapping* (Obr. 25). O editaci v tomto okně více v kap. 7.

Situace 2 - pro daný CM nemáme k dispozici odkaz na sémantický rámec ve FrameNetu

Pokud žádný z CMs synonymní třídy X neobsahuje odkaz na sémantický rámec ve FrameNetu (stává se to výjimečně), inspirujeme se rolemi uvedenými ve FrameNetu u příbuzných sloves. Pokud v seznamu pro výběr role z FrameNetu (Obr. 21 a Tab. 1) nenajdeme vhodnou SR, roli do seznamu bud' doplníme (Obr. 22 – *Add role*, při definování role nezapomeneme zakliknout *From FrameNet* – Obr. 23a), nebo vytvoříme vlastní pojmenování role. Tlačítkem *Add* otevřeme okno pro přidání role (Obr. 22), a poté co se otevře okno *Role definition*, nezaklikneme zde *From FrameNet* (Obr. 23b).

Nové pojmenování role by mělo dostatečně zachycovat význam daného participantu, obdobně jako FEs.⁴⁸ Ale zatímco role z FrameNetu jsou v obecném případě tvořeny na míru jednomu sémantickému rámci, nejdou napříč všemi rámci (např. role *Goal* v jednom rámci vyjadřující *účel*, může v jiném rámci vyjadřovat *místo*), v CzEngClass slovníku směřujeme k tomu, aby SRs měly ve všech třídách, v nichž jsou použity, stejný význam. Nové pojmenování může být jednoslovné (*Transporter*) nebo i víceslovné (*Activity_Event_Proces*). Pokud vytvoříme víceslovné pojmenování SR, zapisujeme pojmenování přes podtržítka (X_Y_Z). Pokud by slova z víceslovného pojmenování mohla existovat jako samostatná pojmenování role (SR: *Topic_Message* x SR: *Topic*, SR: *Message*), píšeme tato slova s velkým písmenem. Tento typ víceslovného pojmenování se používá pro takovou roli, která vyjadřuje obě (nebo více) role z daného pojmenování (říkal o Pavlovi. *Topic*, že lže. *Message* vs. říkal, že Pavel lže. *Topic_Message*).

Pokud některé slovo ve víceslovém pojmenování blíže určuje jiné slovo pojmenování, píšeme toto určující slovo s malým písmenem (SR: *Entity_offered*, SR: *Entity_received*).

7. Jak namapovat členy Rolesetu na členy valenčního rámce

Každý člen synonymní třídy musí vyhovovat definici pro zařazení do příslušné synonymní třídy. Podle této definice jsou v jedné synonymní třídě obsažena jen ta překladová slovesa (slovesné významy), která mají obdobnou významovou i strukturní charakteristiku. Jinými slovy, překladová slovesa jedné synonymní třídy mají stejný nebo podobný, tj. synonymní, význam a stejně nebo obdobně jsou také jejich valenční rámce.

Pro anotaci to znamená, že pro každý CM synonymní třídy X je třeba smysluplně namapovat sémantické role (*Roleset*) obsažené v Rolesetu (kap. 6) třídy X na valenční členy (*Arguments*) valenčního rámce tohoto slovesného členu třídy X.

⁴⁸ Při tvorbě nových pojmenování rolí se můžeme inspirovat také seznamem tematických rolí ve VerbNetu: https://verbs.colorado.edu/verb-index/VerbNet_Guidelines.pdf, i když z teoretického hlediska jde o jiné východisko než ve FrameNetu.

Ideálně by všechny CMs dané třídy měly jen jedno mapování. Tak tomu ale být nemusí. Proto pracujeme s ***Typickým Mapováním Rolesetu a Argumentů*** valenčního rámce (***TypRAM***), které odpovídá většinovému mapování v dané třídě (kap. 6) a charakterizuje významovou a strukturní charakteristiku vybrané synonymní třídy.

Pro mapování rolí z daného *Rolesetu* musí platit, že každá sém. role je na "NĚCO" z valenčního rámce slovesa zařazeného do dané třídy namapována (příklad třídy *vyhodnotit* níže). Tím "NĚCO" může být

- buď člen valenčního rámce (ve třídě *vyhodnotit* černé funktry)
- nebo jiné volné určení, které by u daného slovesa přiřazená SR mohla mít (ve třídě *vyhodnotit* modré funktry)
- anebo zástupný funktor (#any, #sb, #sth), (ve třídě *vyhodnotit* zeleně)

Opačně musí platit, že každý člen valenčního rámce slovesa zařazeného do dané třídy musí být namapován na sémantickou roli z přiřazeného *Rolesetu*.

Z tohoto pravidla existuje jediná výjimka: Pokud je ale v anglickém valenčním rámci některého člena třídy⁴⁹ obsaženo neobligatorní volné doplnění, zapsané jako volné doplnění s otazníkem, při tvorbě *Rolesetu* ho nemusíme (ale můžeme) zohledňovat v mapování, tj. nemusíme (ale můžeme) pro něho mapování doplňovat.

Pokud nějaký z členů valenčního rámce patřícího k potenciálnímu členu dané třídy OPRAVDU nelze namapovat na zvolený *Roleset* příslušné třídy, tento potenciální člen (*kandidát*) do třídy nepatří.

Příklad:

Mapování pro třídu *vyhodnotit*

Roleset			
	<i>Assessor</i>	<i>Phenomenon</i>	<i>Value</i>
<i>Vyhodnotit</i>	ACT	PAT	COMPL
<i>Evaluate</i>	ACT	PAT	COMPL
<i>Rate</i>	ACT	PAT	EFF
<i>Dostat</i>	ORIG	ACT	PAT
<i>Ocenit</i>	ACT	PAT	EFF
<i>Posoudit</i>	ACT	PAT	COMPL
<i>Vyčíslovat</i>	ACT	PAT	EFF
<i>Zhodnotit</i>	ACT	PAT	<u>#sth</u>

⁴⁹ Týká se jen anglických sloves, protože při tvorbě *EngVallexu* platila upravená pravidla pro obsah valenčního rámce: rámec mohl obsahovat i neobligatorní volná doplnění.

Poznámka týkající se *Rolesetu* se v *SynEdu* zapisuje do okna *Note* pod oknem *Roleset*.

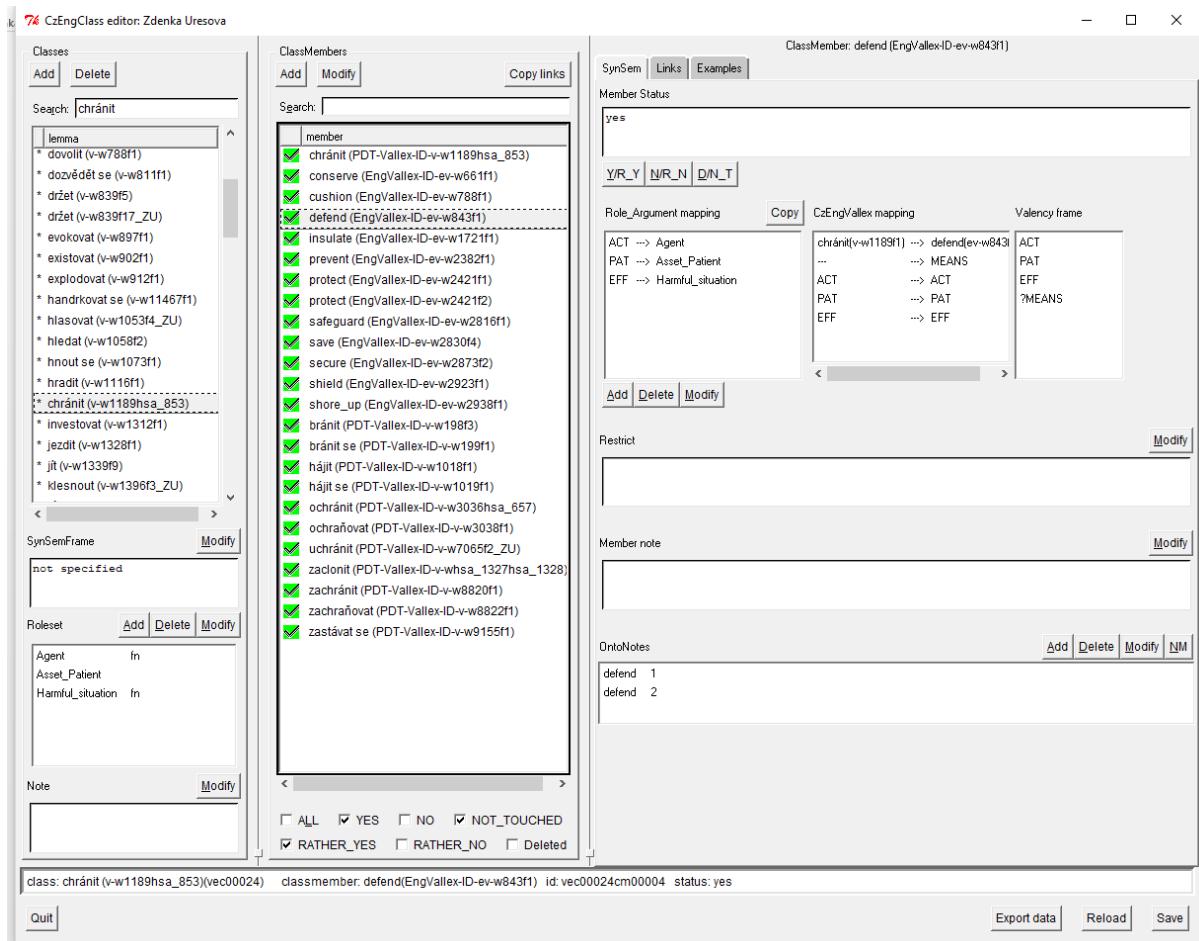
K volbě zástupného funkторu (#any, #sb, #sth) nebo konkrétního funktoru ve valenční teorii FGP.

Pokud si umíme pro daný valenční člen, který potřebujeme namapovat, doplnit z repertoáru valenční teorie FGP nějaký funktor, dáme mu přednost před doplněním zástupných symbolů.

Pro případy, že toto doplnění neumíme, ale daný valenční člen přesto potřebujeme namapovat, protože si ho lze pro dané sloveso představit, doplníme do mapování zástupný funktor pro doplnění **#sb** (životné), nebo **#sth** (neživotné), popřípadě **#any** (obojí).

V *SynEdu* anotujeme toto mapování pod záložkou Syntakticko-sémantické informace (*SynSem*) v okně *Role_Argument mapping* (Obr. 25) následujícím způsobem.

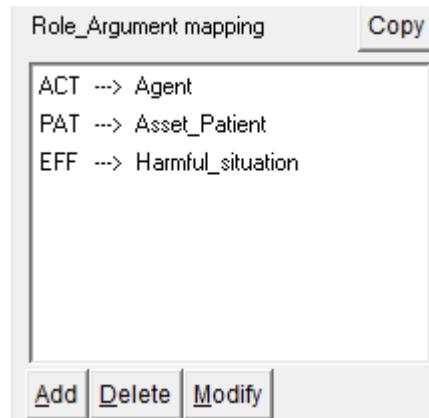
Nejprve otevřeme v *SynEdu* záložku *SynSem* (Obr. 24),



Obr. 24 Záložka *SynSem* s okny: *Role_Argument mapping*, *CzEngVallex mapping* a *Valency frame* pro CM *defend* třídy *chránit* (NeA)

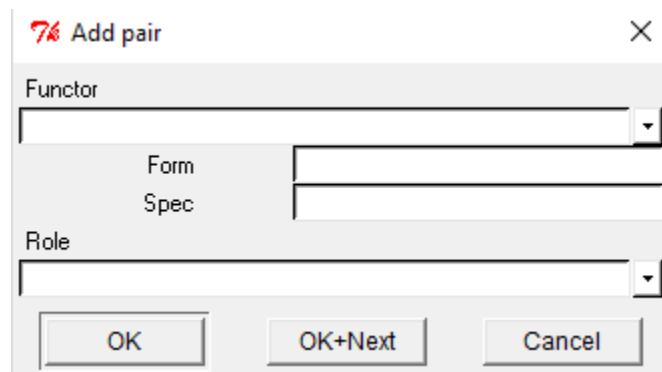
kde je pro každého člena (*defend*) synonymní třídy X (zde např. *chránit*) již zobrazena informace (zprava doleva) o jeho valenčním rámci (VF) - *Valency frame* okno (ACT, PAT, EFF, ?MEANS) a namapování tohoto VF na VF českého slovesa (*chránit*), podle něhož je třída X pojmenována - *CzEngVallex mapping* okno (---> MEANS, ACT > ACT, PAT > PAT, EFF > EFF).

CzEngVallex mapping a *Valency frame* okno nejsou okna editační, informace v nich obsažená má za úkol pomoci s anotací namapování role na argument, kterou zapisujeme do okna v záložce SynSem zcela vlevo: *Role_Argument mapping* (Obr. 25).



Obr. 25 Editační okno *Role_Argument mapping* (CM *defend* třídy *chránit*) - editační tlačítka vlevo dole a tlačítko pro kopírování vpravo nahoře

Pokud jsme si pro danou třídu už připravili *Roleset*, je situace jednoduchá. Pomocí *Add* otevřeme okno pro mapování (Obr. 26) a postupně přes tlačítko *OK+Next* namapujeme valenční členy daného CM synonymní třídy na připravené role.



Obr. 26 Editační okno pro mapování argumentu a role

7.1. K realizaci sémantické role: mapování rolí

Ve většině případů platí, že jedna SR se mapuje na jeden argument z valenčního rámce členu třídy.

Může ale nastat situace,

1. kdy je jedna SR sice realizována jako jeden argument, který je ale ve valenčním rámci zapsán jako několik alternujících funktorů, viz 7.1.1.
2. kdy je jedna SR realizována dvojím způsobem (nebo i více způsoby), jednou (přímo) jako argument valenčního rámce slovesa-členu třídy a jindy jako tzv. „vnořený“ argument, viz 7.1.2. Vnořeným argumentem rozumíme argument ke slovesu, které je kořenem vedlejší věty reprezentující některý přímý argument z valenčního rámce členu třídy.
3. kdy je jedna SR realizována několika argumenty z (jednoho) valenčního rámce členu třídy, viz 7.1.3.

7.1.1. Mapování sémantické role na alternující funktoře

Pro zápis mapování sémantické role na alternující funktoře vybereme mezi funktoře z polička Functor (Obr. 26) zástupný symbol `#alt` a do polička *Spec* (Obr. 26) zapíšeme všechny alternující funktoře v rámci. Oddělujeme je čárkou (POZOR: nepíšeme sem hranaté závorky).

Příklad:

Ve třídě *držet* (*v-w839f17_ZU*) je zařazen člen *udržet* s valenčním rámcem obsahujícím alternující funktoře LOC a MANN.

udržet⁵_{3x,23x} **ACT(1)** **PAT(4)** **LOC(=)|MANN(=)**

udržíme byt pohromadě.MANN(=); u. je naživu.MANN(=); u. ho ve vzorném pořádku.LOC(=);
udržet děti v tichu.LOC(=); u. elektrárnu v provozu.LOC(=); u. syna v nevědomosti.LOC(=)

Při mapování alternujících funktořů vybereme z nabídky funktořů zástupný symbol `#alt` a do pole *Spec* zapíšeme alternující funktoře "LOC,MANN". V SynEdu se pak objeví toto mapování:

`#alt [LOC,MANN]->STATE`

Pokud by bylo v rámci více alternujících funktořů, zapíšeme je do pole *Spec* všechny. Například: MANN,ACMP,CPR.

7.1.2. Mapování sémantické role realizované jinak než přímým valenčním členem

Pro zápis mapování sémantické role realizované **přímo jako argument valenčního rámce členu třídy** použijeme okno pro mapování (Obr. 26) a SR namapujeme na odpovídající valenční člen z rámce daného členu synonymní třídy.

Pro zápis mapování sémantické role **realizované jako argument, který je vnořený do jiného „přímého“ argumentu** z valenčního rámce členu třídy se řídíme následujícími pokyny.

Zápis je formalizovaný takto:

- zápis **FUNC1[FUNC2]** označuje, že role se namapuje na vnořený argument s FUNC2, který visí na valenčním argumentu slovesa s FUNC1,
- popřípadě zápis **FUNC1[FUNC2[FUNC3]]** označuje, že role se namapuje na vnořený argument s FUNC3, který visí jako potomek na FUNC2 a FUNC2 visí na valenčním argumentu slovesa s FUNC1.

Pro zápis FUNC1[FUNC2] se FUNC2 zapíše do pole *Spec* bez závorek, editor si je doplní.

Pro zápis FUNC1[FUNC2[FUNC3]] se do pole *Spec* musí použít závorky: FUNC2[FUNC3], ve výsledku pak editor jedno závorkování doplní.

Příklad:

Zápis **PAT[ACT]** u třídy *dovolit* (*v-w788f1*), člen *allow* (*ev-w86f4*) indikuje, že role *Protagonist* je "vnořeným" **ACTorem**, který visí na PATientu slovesa *allow*.

The Food and Drug Administration.ACT allowed the company.ACT to begin.PAT arketing a new lens for use in cataract patients.

Příklad:

Zápis **ACT[RSTR[PAT]]** - u třídy *dít se* (*v-w492f2*), člen *be* (*ev-w218f3*) indikuje, že role *Entity* je "vnořeným" **PATientem**, který visí na RSTR, které visí na ACT slovesa *be*.

There is nothing.ACT wrong.RSTR with the economy.PAT.

7.1.3. Mapování jedné sémantické role na několik argumentů jednoho valenčního rámce současně.

Sémantická role může být realizována i **několika argumenty** z (jednoho) valenčního rámce členu třídy **současně**, jde o tzv. **rozpad jedné role** na dvě valenční doplnění.

Například některé členy ve třídě *stěžovat si* (*v-w652f1*), mají v rámci čtyři argumenty - ACT, ADDR, PAT, EFF. Zde je jedna sémantická role (Complaint) realizována společně (dohromady) dvěma argumenty (PAT a EFF):

Pro realizaci **jedné role dvěma argumenty** (PAT a EFF) platí mapování:

Výbor.ACT si jim.ADDR stěžuje na obchody.PAT, že neplní.EFF svou povinnost.

ACT > Complainier

ADDR > Addressee

PAT > Complaint

EFF > Complaint

Pro zápis mapování jedné sémantické role **na jeden argument** nepotřebujeme zvláštní pokyny, neboť přímočaré realizaci sémantické role odpovídá přímočaré mapování na jeden argument.

Sloveso *complain* ze třídy *stěžovat si* (*v-w652IfI*) má v rámci na rozdíl od jiných jen tři argumenty - ACT, ADDR, PAT.

Pro realizaci **jedné role jedním argumentem** (PAT) platí mapování:

Foreigners.ACT complain to Washington.ADDR that they have.PAT limited access.

ACT > Complainier

ADDR > Addressee

PAT > Complaint

Tzv. rozpad role je typický například pro slovesa dicendi:

Třída *soudit* (*v-w6270fI*) má přiřazen roleset Cognizer Phenomenon.

Některé členy v této třídě, např. samo sloveso *soudit*, mají v rámci ACT, PAT, EFF. Tomu odpovídá mapování **jedné role na dva argumenty**:

O exportu.PAT se soudí, že v srpnu silně vzroste.EFF

ACT > Cognizer

PAT > Phenomenon

EFF > Phenomenon

V téže třídě může být jedna role (Phenomenon) realizována i jen jedním argumentem:

Nicméně pokud ministr.ACT usoudí, že transakce učiní.PAT dopravce neschopným provozu

ACT > Cognizer

PAT > Phenomenon

7.2. Specifikace funktoru při mapování na sémantickou roli

Někdy je třeba *funktor* namapovaný na *sémantickou roli* blíže specifikovat. Specifikaci zapisujeme do políčka *Spec* (Obr. 26), nemá zatím nijak formálně definovanou formu, použijte volný zápis, ale pokud možno jednoznačný. Při zápisu do pole *Spec* nepíšete hranaté závorky. Doplní se automaticky.

Příklad:

Sloveso *vzepřít se* má identický valenční rámec (*funktory* ACT, PAT) pro dva významy.

- Pro význam *vzepřít se něčemu neživotnému* (*vzepřít se její vůli, vzepřít se jeho rozkazům*) zařazenému do třídy *handrkovat se* je doplněn *Roleset*: Arguer1, Arguer2, Issue. Mapování by zde mělo zachytit přivlastkové vyjádření *sémantické role* Arguer2.

ACT -> Arguer1

#sb[poss.pronoun/adj.] -> Arguer2

PAT -> Issue

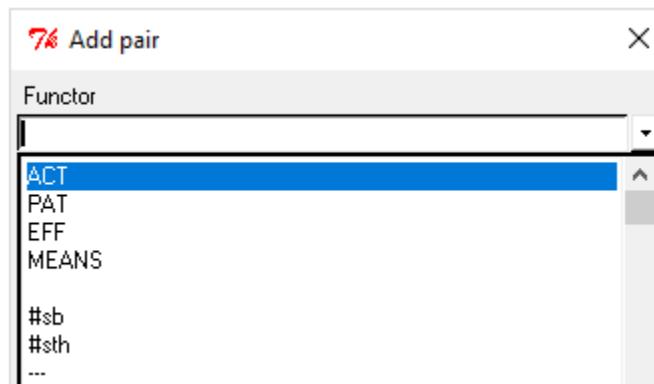
2. Pro význam *vzepřít se někomu* (*vzepřít se nepříteli*) další specifikace není potřeba, doplněný *Roleset*: Arguer1, Arguer2 je namapován bez ní.

Mapování

ACT -> Arguer1

PAT -> Arguer2

Valenční člen zapisujeme jako *Functor*, který vybereme z předpřipravené nabídky funktorů rozkliknutím šipky na konci políčka pro funkтор (Obr. 27).



Obr. 27 Anotace funktoru

V nabídce se vždy objevují nejprve ty funktry, které jsou obsaženy ve VF daného slovesa (pro *defend*: ACT, PAT, EFF, MEANS). Následují dva zástupné funktry (#any, #sb, #sth) pro případ, že je třeba namapovat roli z daného *Rolesetu* na funktor, který sice není zapsán ve VF slovesa, které anotujeme, ale lze si jej pro dané sloveso představit. Absence valenčního členu ve VF může být z různých důvodů – chyba v rámci, konvence pro zápis VF⁵⁰ apod.

Příkladem je namapování role *Harmful_situation* na zástupný funktor #sth u slovesa *conserve* třídy *chránit* (Obr. 28), jehož VF neobsahuje EFF, který je na roli *Harmful_situation* namapován

⁵⁰ Ve valenční teorii Funkčně generativního popisu (FGP) jazyka se do VF zapisují jen obligatorní (ACT, PAT, ADDR, EFF, ORIG) a fakultativní (otazník před funktem, např. ?ACT) aktanty a obligatorní volná doplnění (LOC). To platí pro PDT-Vallex; v EngVallexu jsou do VF zapisovány s otazníkem i volná doplnění.

u většiny členů třídy chránit.

Obr. 28 Mapování role na zástupný funkтор pro *conserve*

Sémantickou roli zapisujeme do políčka *Role*, kterou vybereme z předpřipravené nabídky SRs rozkliknutím šipky na konci políčka pro roli (Obr. 28).

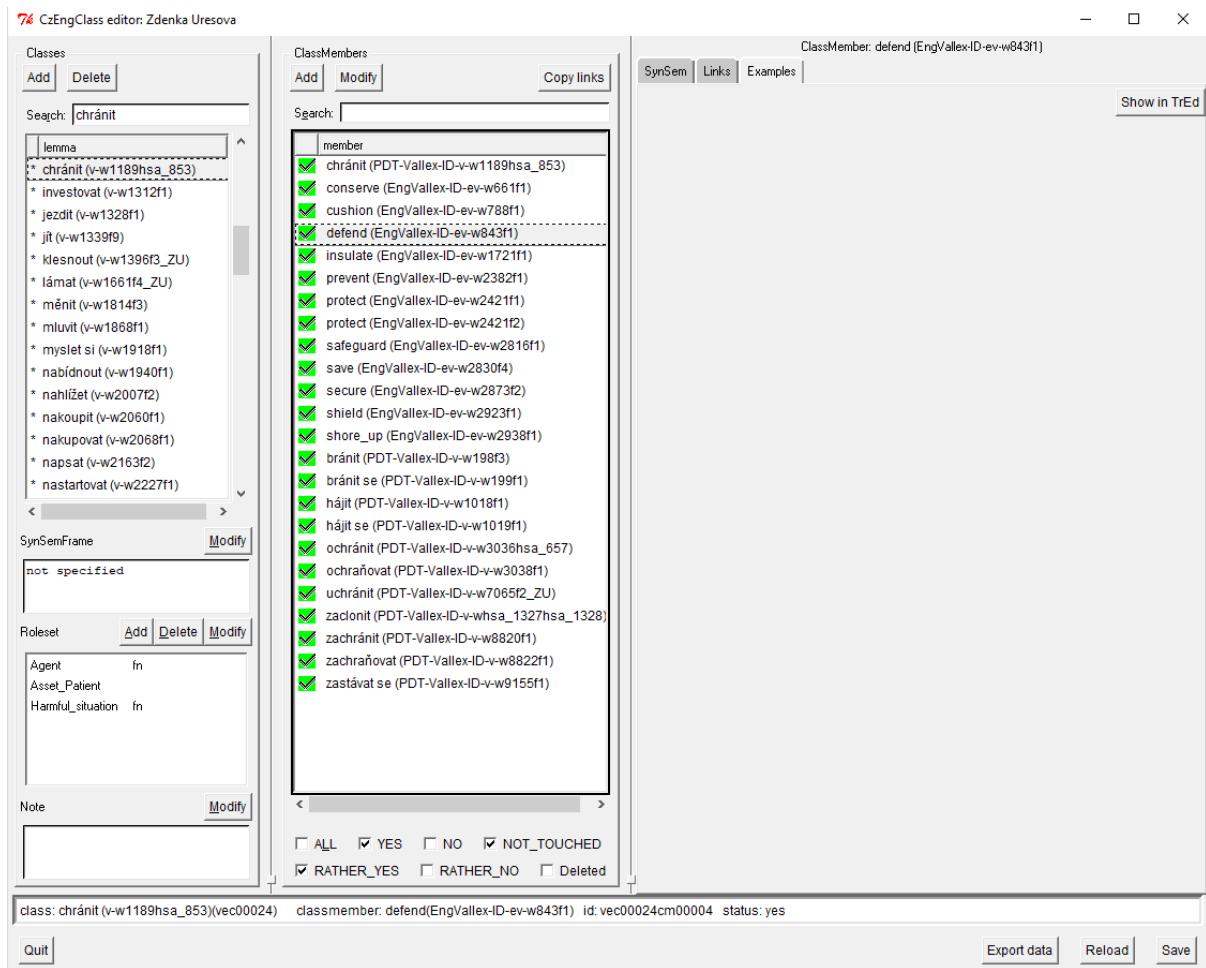
Obr. 28 Anotace sémantické role

V nabídce se vždy objevují nejprve ty role, které jsou obsaženy v *Rolesetu* daného slovesa (pro *defend*: Agent, Asset_Patient, Harmful_situation).

Pokud ale *Roleset* pro danou synonymní třídu ještě nemáme vytvořený (kap. 6.2), postupujeme následovně. Prohlédneme si odkazy do FrameNetu u všech CMs dané třídy, tím zjistíme, jaký

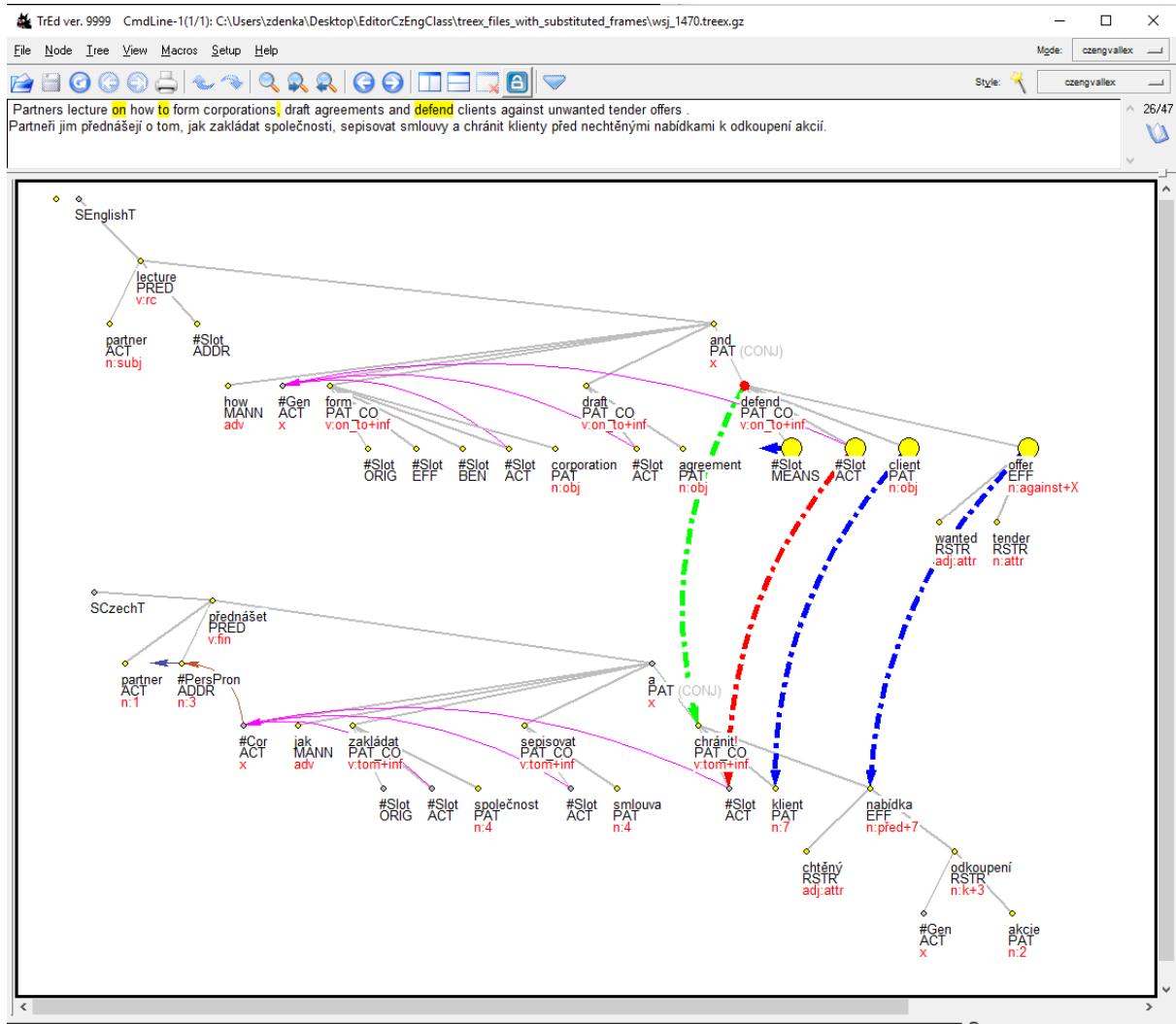
sémantický rámec z FrameNetu se v odkazech objevuje nejčastěji a prohlédneme si jeho FEs, zejména *Core FEs*. Pokud máme u jednotlivých CMs v odkazech do FrameNetu i další sémantické rámce, zkонтrolujeme FEs i u nich. Vytvoříme jejich přehledný seznam a zvážíme, které vybereme jako členy našeho *Rolesetu*. Pokud nám pojmenování nabízených SRs (FEs) z FrameNetu nevyhovuje, pojmenování buď upravíme, nebo vymyslíme nová. Nově vytvořený *Roleset* pak namapujeme na argumenty VF jednotlivých CMs dané třídy, jak je popsáno výše.

Při mapování rolí a argumentů nám, stejně jako při anotování příslušnosti člena do synonymní třídy (*Member Status*), mohou pomoci příklady z korpusu PCEDT. Zatím nejsou⁵¹ v *SynEdu* implementovány. Prohlížíme si je tak, že otevřeme překladové páry sloves v anotačním nástroji *TrEd* (*Tree Editor*) v záložce *Examples* přes tlačítko *Show in TrEd* (Obr. 29 a 30).



Obr: 29 *SynEd* - Záložka *Example* s tlačítkem *Show in TrEd* (NeA)

⁵¹ POZOR Změna, viz Rozšíření *SynEdu*, kap. 8.2.

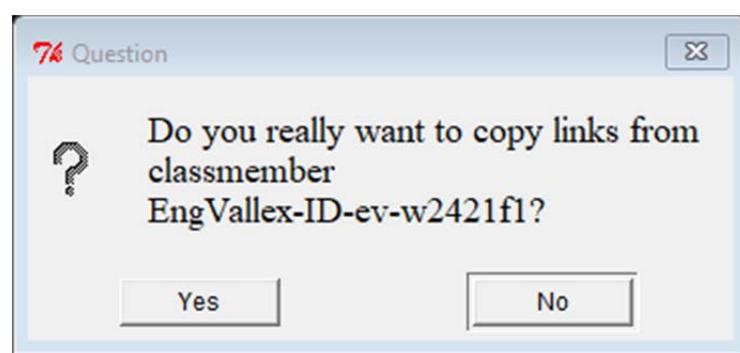


Obr. 30 Příklad pro překladový pár *defend – chránit* v anotaci PCEDT, zobrazený v TrEdu

Pro usnadnění anotace jsou v *SynEdu* implementována tlačítka pro kopírování (Obr. 31 a 33). Lze je využít, pokud se odkazy (Obr. 31) nebo mapování role na argumenty (Obr. 33) u jednoho nebo více CMs dané třídy shodují.

Chceme-li zkopirovat odkazy zapsané u jednoho CM, stiskneme tlačítko *Copy links* (Obr. 31), když “stojíme” v okně *ClassMembers* na CM, jehož odkazy chceme kopírovat, a potvrďme (Obr. 32), že chceme zkopirování opravdu provést. Odkazy se automaticky zkopiují do všech CMs se stejným lemmatem v dané třídě (třída *chránit* – při stisknutí *Copy links* na CM *protect-f1* se odkazy automaticky zkopiují do *protect-f2*).

Obr. 31 Kopírování odkazů z jednoho CM k jinému CM téže třídy



Obr. 32 Potvrzení kopírování odkazů

Obdobně postupujeme, pokud se shoduje mapování role na argument u jednoho nebo více CMs dané třídy, s tím rozdílem, že chceme-li zkopirovat mapování zapsané u jednoho CM, stiskneme

tlačítko *Copy* (Obr. 33), které zčervená; pak klikneme na CM, kam chceme mapování zkopiřovat, a potvrdíme, že chceme tuto operaci provést.

Pokud kopírujeme mapování do CM, kde je mapování už vyplněno, editor nabídne možnost *Merge* a *Replace*. Pokynem *Replace* se stávající mapování u nového CM nahradí mapováním, které chceme zkopiřovat, zatímco pokyn *Merge* přidá z kopírovaného zdrojového mapování pouze ty dvojice (*Functor* → *Role*), které stávající mapování v novém CM ještě neobsahovalo.



Obr. 33 Kopírování mapování role a argumentu z jednoho CM třídy k jinému CM téže třídy

8. Anotační nástroje pro tvorbu slovníku CzEngClass

Při vytváření slovníku CzEngClass používáme dva základní anotační formáty, **Excel** (kap. 3.1) a speciální anotační editor *SynEd* (kap. 8.1.). Zatímco Excel využíváme zejména pro anotaci statusu členů třídy (zařazení do třídy), *SynEd* editor je určen pro celkovou anotaci.

8.1. Synonyms Editor

Anotační nástroj pro tvorbu slovníku CzEngClass, nazvaný ***Synonyms Editor*** (*SynEd*), byl tvořen „od nuly“ v průběhu prvního roku projektu. Snažili jsme se, aby co nejvíce zjednodušoval práci anotátora a zároveň umožňoval převod vytvořených informací do internetové podoby slovníku, který plánujeme poslední rok projektu zpřístupnit online jako součást služeb v projektu LINDAT/CLARIN.⁵²

Jsme si vědomi toho, že *SynEd* dosud nemá finální podobu. Některé věci dosud nejsou implementovány, jiné se mění na základě požadavků anotátorů. I přesto doufáme, že se vám

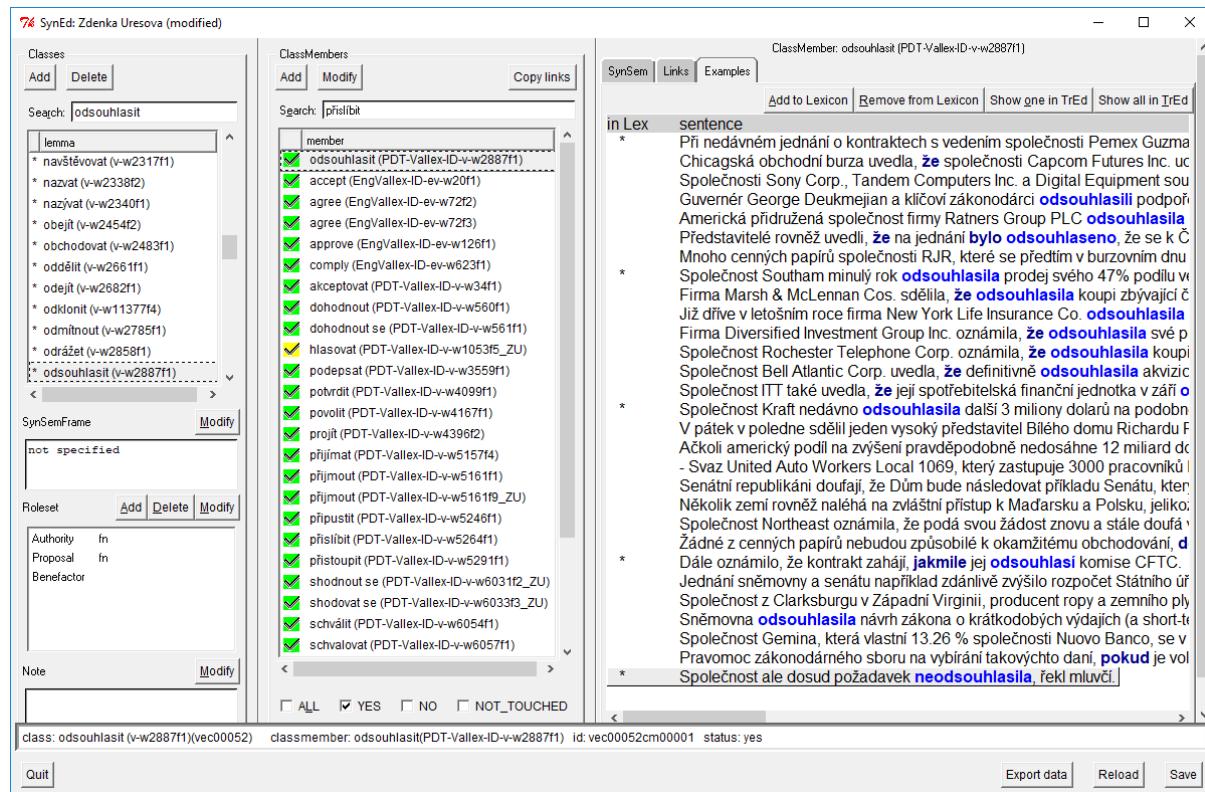
⁵² <https://lindat.mff.cuni.cz/cs/>

s editorem bude dobře pracovat a usnadní vám práci na slovníku CzEngClass. Za veškeré připomínky k editoru budeme vděčné.

Pokyny k instalaci *SynEdu* jsou součástí „technického“ manuálu, který budete mít k dispozici. Pokyny, jak *SynEd* používat, jsou popsány výše. Při anotování je důležité práci průběžně ukládat a komitovat do SVN systému (kap. 9), aby se anotace neztratily.

8.2. Rozšíření editoru o přiřazování příkladových vět

Nově jsou v editoru doplněny příkladové věty (Obr. 34) a tlačítka *Add to Lexicon*, *Remove from Lexicon*, *Show one in TrEd*, *Show all in TrEd*. Z vět vybíráme a vkládáme do slovníku ty nejvíce reprezentativní, max. 5 vět.



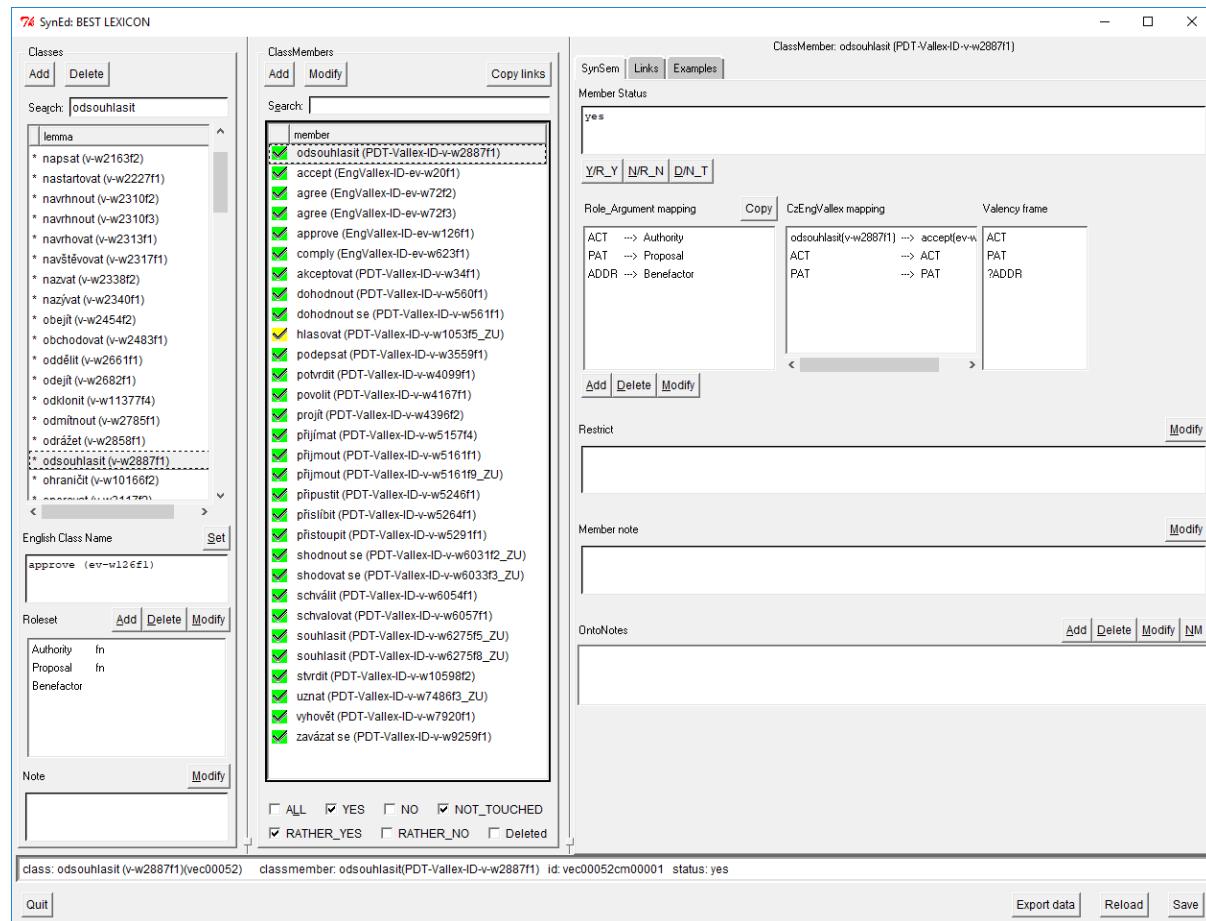
Obr. 34 *SynEd* - Záložka *Example* s tlačítky *Add to Lexicon*, *Remove from Lexicon*, *Show one in TrEd*, *Show all in TrEd* (NeA)

8.3. Rozšíření editoru o přiřazování anglického slovesa do názvu třídy

Název synonymní třídy měl po dlouhou dobu anotování dat prozatím název reprezentovaný českým slovesným významem a identifikací valenčního rámce (kap. 3.1.) - např. sloveso *prozkoumat* (*v-w4633f1*).

Aktualizace editoru nyní umožňuje přiřadit do názvu třídy i reprezentativní anglické sloveso. Zatímco české sloveso v názvu třídy nijak neanotujeme, je automaticky předvybráno, anglické sloveso vybíráme z přiřazených anglických členů třídy. Postupujeme přitom tak, že po jeho označení klikem v okně *ClassMembers* klikneme na tlačítko *Set* v okně *English Class Name* (okno původně označené *SyneSemFrame*, vlevo pod oknem *Classes*, Obr. 35) a vybrané anglické sloveso se sem automaticky zapíše.

Za reprezentativní anglický člen třídy považujeme takové anglické sloveso, které má **Typické Mapování Rolesetu a Argumentů** valenčního rámce (kap. 6.) a na základě svých sémantických a strukturních vlastností je, stejně jako reprezentativní české sloveso v názvu třídy, typickým zástupcem anglických členů třídy.



Obr: 35 *SynEd* – přiřazení anglického slovesa do názvu třídy – okno *English Class Name*

9. Archivace a distribuce anotovaných dat

Pro správu (distribuci a archivaci) vašich anotovaných dat byl pro každého anotátora zřízen speciální **Subversion (SVN)** repositář, nazveme ho **ÚFAL-repositář**, na adrese:

<https://svn.ms.mff.cuni.cz/svn/CzEngClass/data/anotace/<vaše iniciály>>

Pokyny k instalaci klienta TortoiseSVN („svnky“) jsou součástí „technického“ manuálu, který budete mít k dispozici.

Pro práci anotátora potřebujete umět následující operace:

Checkout - založení tzv. *pracovní kopie repositáře* na vašem počítači (SVN klient prvotně zkopiuje data z repositáře na ÚFAlu do vašeho počítače a začne sledovat změny).
Tuto operaci budete potřebovat jen jednou na začátku práce.

Commit - promítnutí změn provedených ve vaší pracovní kopii repositáře do ÚFAL-repositáře (SVN klient zkopiuje data z vaší pracovní kopie zpět do ÚFAL-repositáře – sám pozná, která se změnila a zkopiuje pouze ty).

Tuto operaci budete potřebovat pravidelně pro odevzdávání anotovaných dat.

Update - promítnutí změn z ÚFAL- repositáře do vaší pracovní kopie (SVN klient zkopiuje nová data a změny z ÚFAL- repositáře do vaší pracovní kopie, aby se obsah pracovní kopie repositáře a ÚFAL- repositáře shodoval)

Tuto operaci budete potřebovat pravidelně pro získávání nově přidělené dávky ostrých dat z ÚFAL-repositáře.

Revision - číslo aktuální „revize“ (verze) obsahu ÚFAL- repositáře, každý Commit ho zvýší (kompletní historie změn v repositáři se eviduje, ke starší verzi se můžeme kdykoliv vrátit)

Merge - Systém SVN je určen pro spolupráci více lidí nad stejnými soubory (v jednom okamžiku existuje několik pracovních kopií). Může tedy občas dojít ke konfliktu, potřebě sloučit změny z ÚFAL- repositáře se změnami ve vaší pracovní kopii repositáře. Pro účely našeho projektu má však každý anotátor přidělen v ÚFAL-repositáři vlastní podadresář, do kterého ostatní anotátoři nemají přístup. Konflikty tedy běžně řešit nebudeme.

PART III – Technical guidelines

10. Installation of TrEd for Windows

Download Setup Package

From <https://ufal.mff.cuni.cz/tred>, from the Download section download the file

tred-Installer-perl-included.exe

This file contains TrEd, Strawberry Perl and required Modules.

Windows

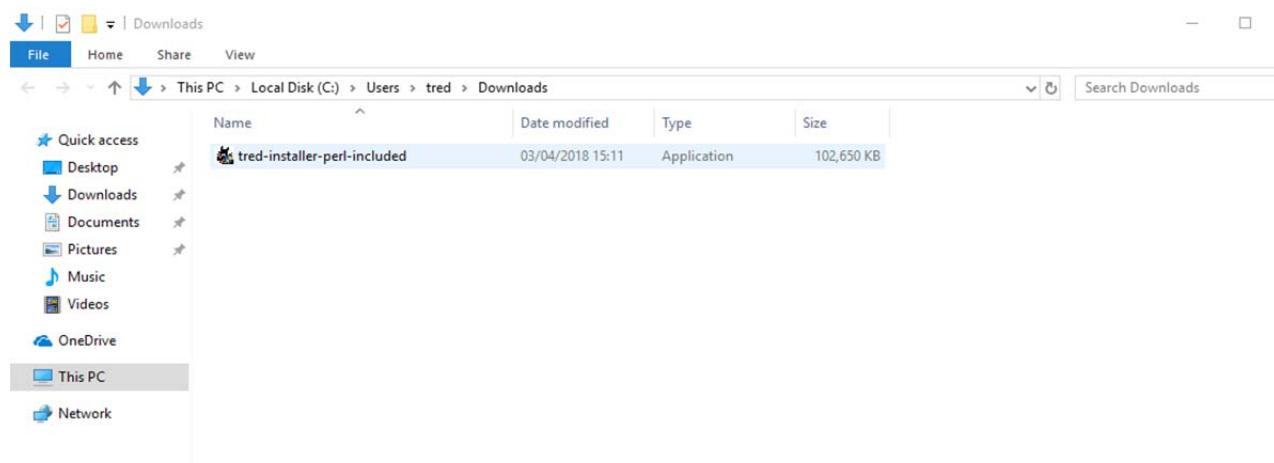
tred-Installer-perl-included.exe (108M) - TrEd + Modules + Strawberry Perl

This is a full Installation package of TrEd for Microsoft Windows. This Installer contains TrEd, required Perl Modules and a recent version of Strawberry Perl.

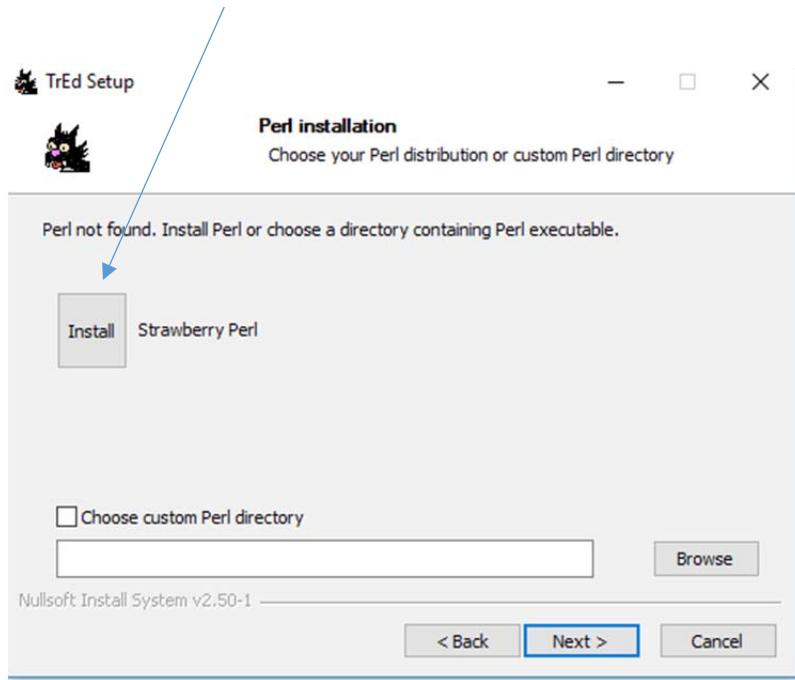
tred-Installer.exe (30M) - TrEd + Modules

10.1 How to install TrEd

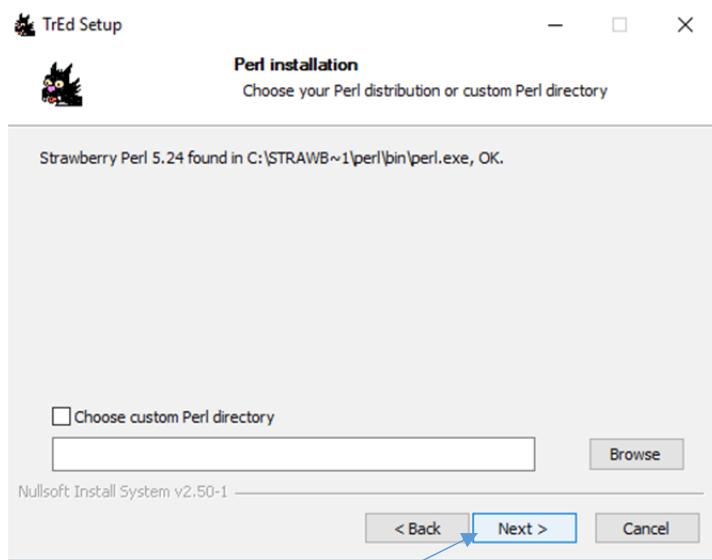
Start the Installation by running the downloaded file



If you don't have Perl Installed, the TrEd Installer will offer to Install it - you can confirm this by pressing the **Install** button.



After Installing Perl (or if we have the Perl Installed from before) the following window will be shown:



Using the **Next** button, you continue with the Installation of TrEd itself.

After Installing TrEd, you must edit the **tred.bat** file in the **tred** directory - edit the last but one line by adding

~~%TRED_DIR%~~

in front of *tred*. Then the edited line will look like this (see the following picture- the line in red):

...\\perl.exe %TRED_DIR%\\tred %*

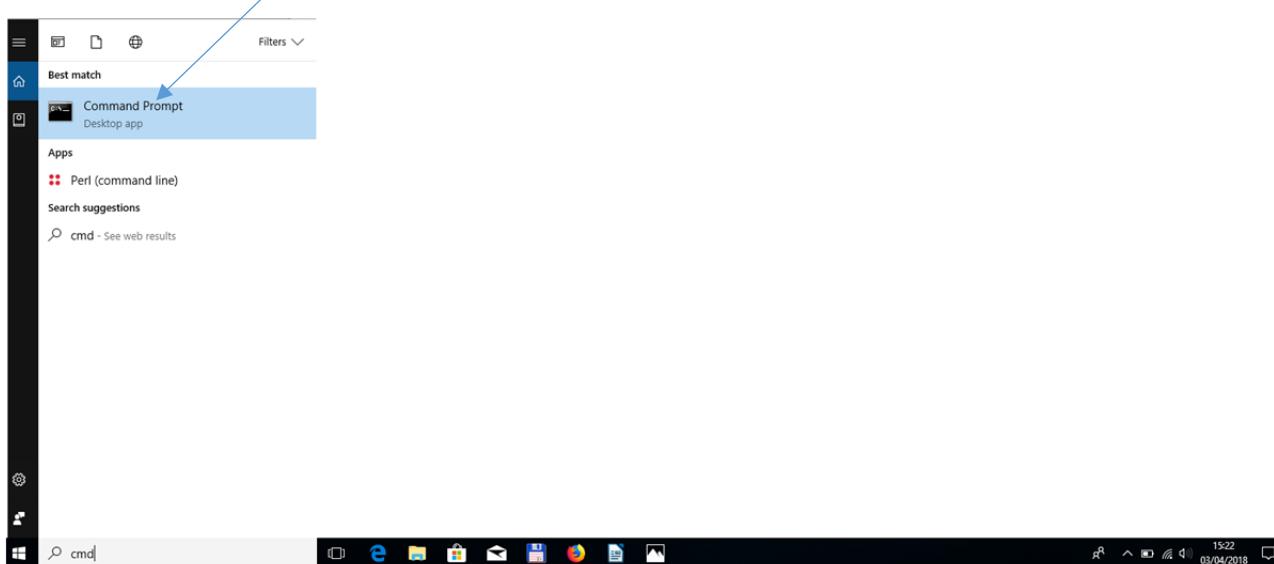


```
tred - Notepad
File Edit Format View Help
@echo off
set PATH=C:\Users\tred\tred\c\bin;C:\Users\tred\tred\bin;C:\Users\tred\tred\dependencies\bin;%PATH%
set PERL5LIB=C:\Users\tred\tred\dependencies\lib\perl5;C:\Users\tred\tred\dependencies\lib\perl5\MSWin32-x86-multi-thread-64int;%PERL5LIB%
set TRED_DIR=C:\Users\tred\tred

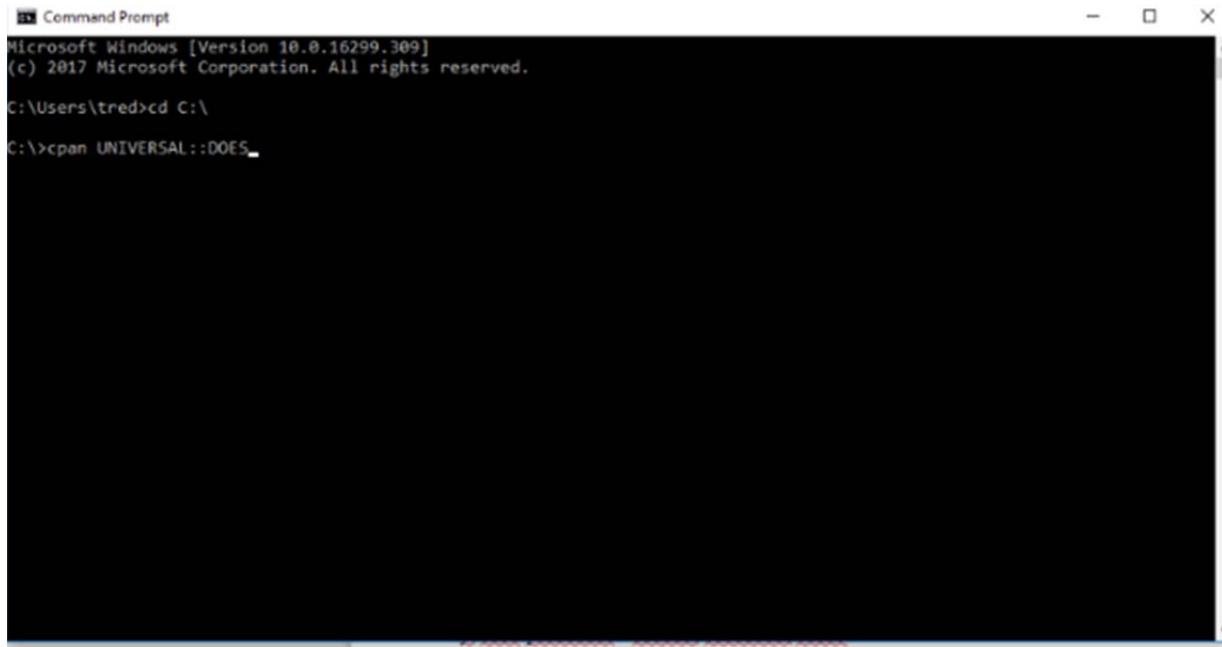
if "%OS%" == "Windows_NT" goto WinNT
C:\\STRAWB~1\\perl\\bin\\perl.exe tred %1 %2 %3 %4 %5 %6 %7 %8 %9
goto end
:WinNT
"C:\\STRAWB~1\\perl\\bin\\perl.exe" "%TRED_DIR%\\tred" %*
:end
```

10.2. Installation of additional Perl modules

First you run **Command Prompt**:



Next you move to the "c:\\" directory (command "cd c:\\") and then use the command "**cpan module_name**" to Install required Perl modules (see their list below).



```
Command Prompt
Microsoft Windows [Version 10.0.16299.309]
(c) 2017 Microsoft Corporation. All rights reserved.

C:\Users\tred>cd C:\

C:\>cpan UNIVERSAL::DOES
```

The list of Perl modules you need to Install:

UNIVERSAL::DOES
 Readonly
 Tk
 Treex::PML::Schema::CDATA
 XML::LibXML::Iterator
 URI::Encode
 Tie::IxHash

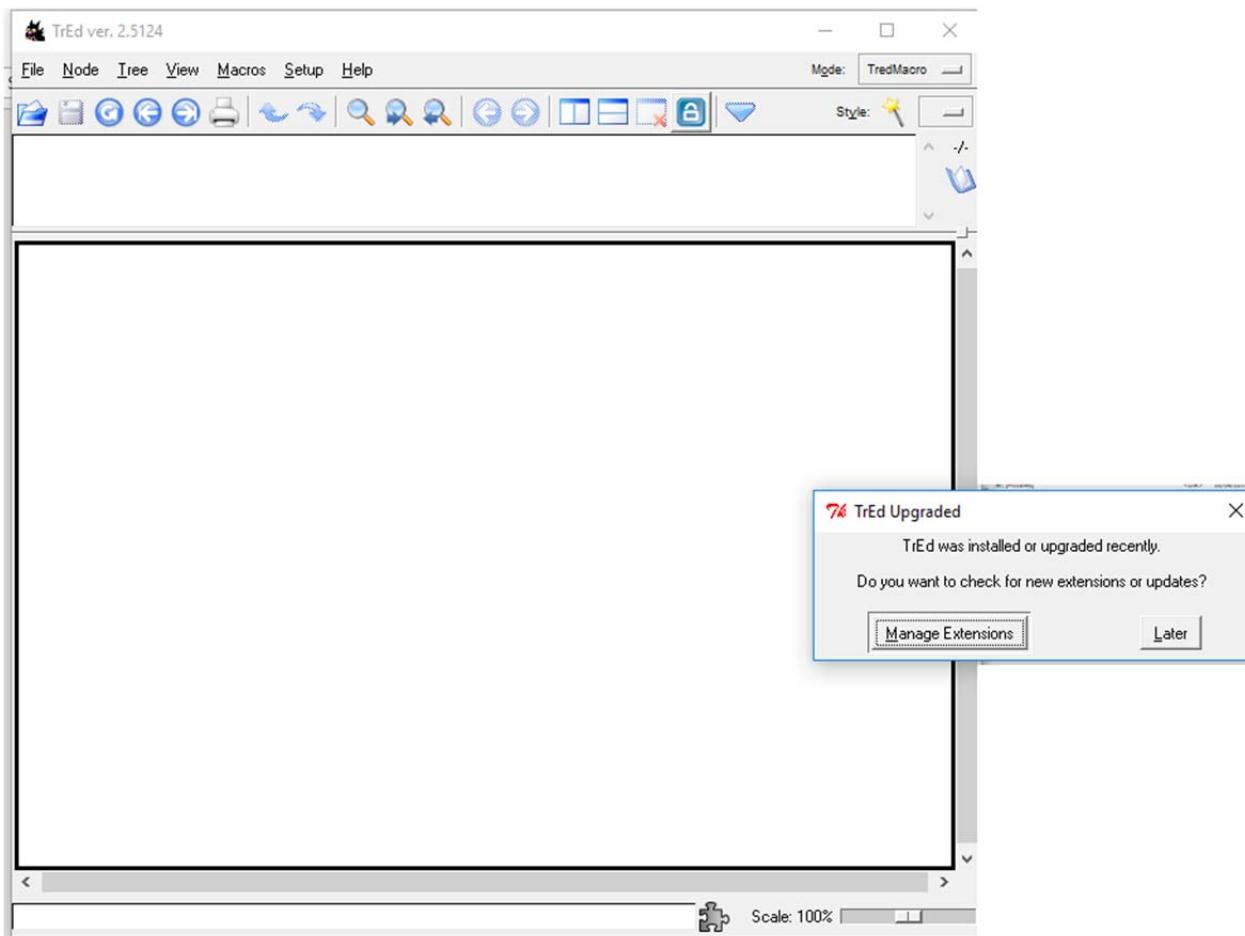
10.3. How to Run TrEd and Install TrEd's Extensions

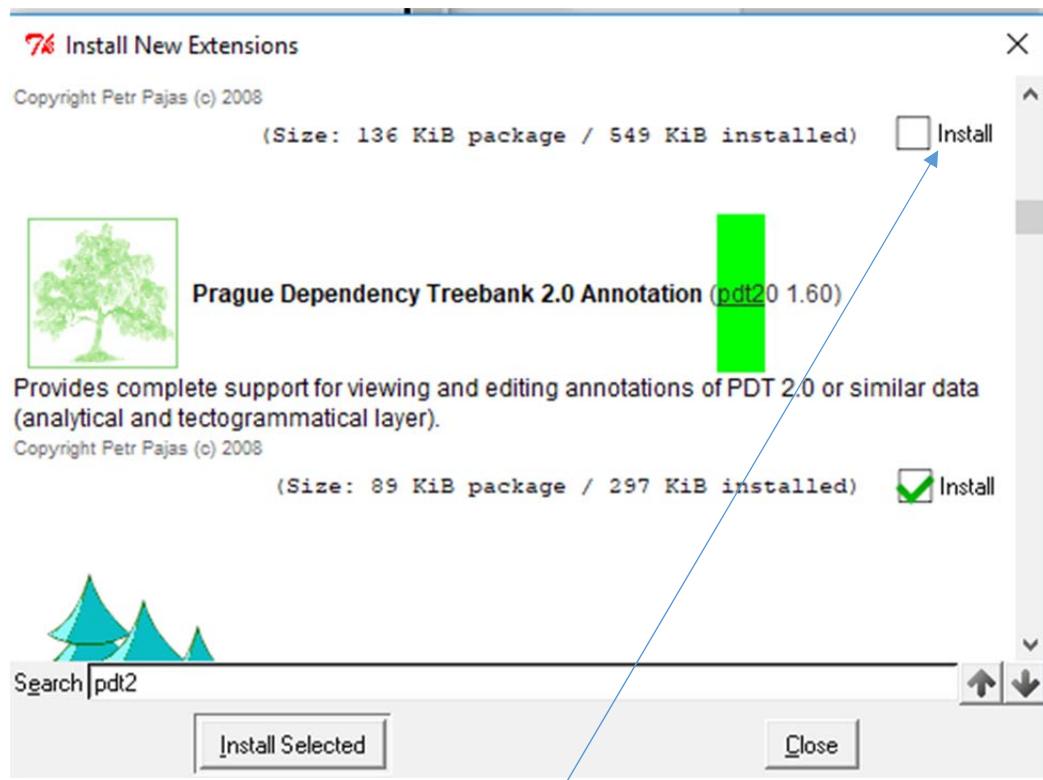
You start to run TrEd by using the *tred.bat* file (edited in section 1.2) in the *tred* directory. If it does not work, you run, similar to the previous step, the **Command Prompt**, where you move to the *tred* directory from where you run *tred.bat*. Any error messages that appear during startup should be sent to Eva Fučíková: *fucikova@ufal.mff.cuni.cz*.

If you succeeded in running TrEd, you continue with adding TrEd extensions (so-called modules) to work with our data. When TrEd opens, the **TrEd Upgraded** menu appears. In this menu you press the **Manage Extensions** button.

In case of some problems, it is also possible to add TrEd extensions by choosing the **Setup→ManageExtension** in the TrEd **menu**, which opens the **ManageExtensions** window where you press the **GetNewExtensions** button.

Both of the above described procedures lead to the opening of the **Install New Extensions** window, where you find the necessary extensions. At the bottom window **Search**, you search for the following four extensions: **pdt2.0**, **pdt-vallex**, **pedt**, and **czengvallex**.





For these extensions, check the **Install** on the right hand side.

If all the necessary extensions are selected, you press the **Install Selected** button.

11. Installation of TortoiseSVN

From <https://tortoisेस्वन.net/downloads.html> you download the TortoiseSVN version you need - (32/64 bit).

The screenshot shows the TortoiseSVN download page. At the top, there's a navigation bar with links for Home, About, Downloads, Translations, Support/Docs, Other tools, and Support this project. Below the navigation is the Tortoise SVN logo. The main content area has two columns: 'Info' on the left and 'Downloads' on the right. The 'Info' column contains links for About, Screenshots, Testimonials, News Archive, Support, FAQ, Help files, Useful tips, Translations, and Misc. The 'Downloads' column displays the current version as 1.12.0. It provides download links for TortoiseSVN 1.12.0 (32-bit and 64-bit) and TortoiseSVN 1.10.3 (32-bit and 64-bit). A note at the bottom explains context menu issues for x64 OS users.

Next, you run the **TortoiseSVN....exe** file:

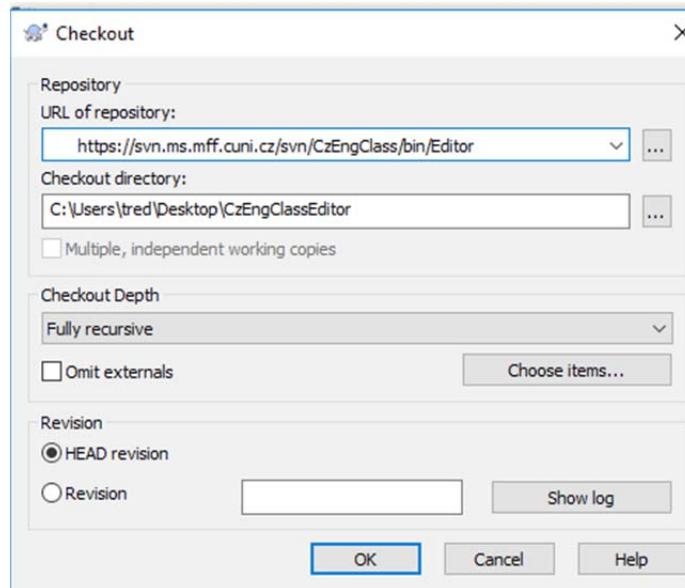
The screenshot shows the Windows File Explorer interface. The left sidebar shows 'This PC' with icons for Quick access, Desktop, Downloads, Documents, Pictures, Music, Screenshots, Videos, OneDrive, This PC, and Network. The main pane displays the 'Downloads' folder, which contains two items: 'TortoiseSVN-1.9.7.27907-x64-svn-1.9.7' (Windows Installer, 17,172 KB) and 'tred-installer-perl-included' (Application, 102,650 KB). The status bar at the bottom indicates '2 items 1 item selected 16.7 MB'.

12. Installation of CzEngClass Editor

12.1. Editor

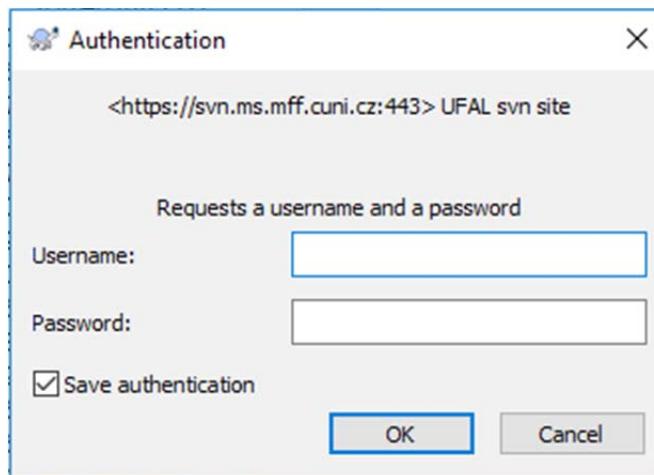
You create a new ***EditorCzEngClass*** directory at the desktop (possibly at a different location, it is up to you). Next, use right-click on this directory and select ***SVN Checkout*** from the menu. A ***Checkout*** window opens; here you fill into the ***URL of repository*** line the following addressee

<https://svn.ms.mff.cuni.cz/svn/CzEngClass/bin/Editor>



(In the above picture, replace tred in ...Users\tred\... by your Windows username)

When you press the ***OK*** button, you will be asked for credentials (they will be mailed to you).



After downloading the Editor from **svn**, you go to the ***EditorCzEngClass/resources/tred_filelists*** directory and extract the ***filelist.zip*** here. This way you get a lot of files with ".fl" ending, placed in the ***EditorCzEngClass/resources/tred_filelists/*** directory.

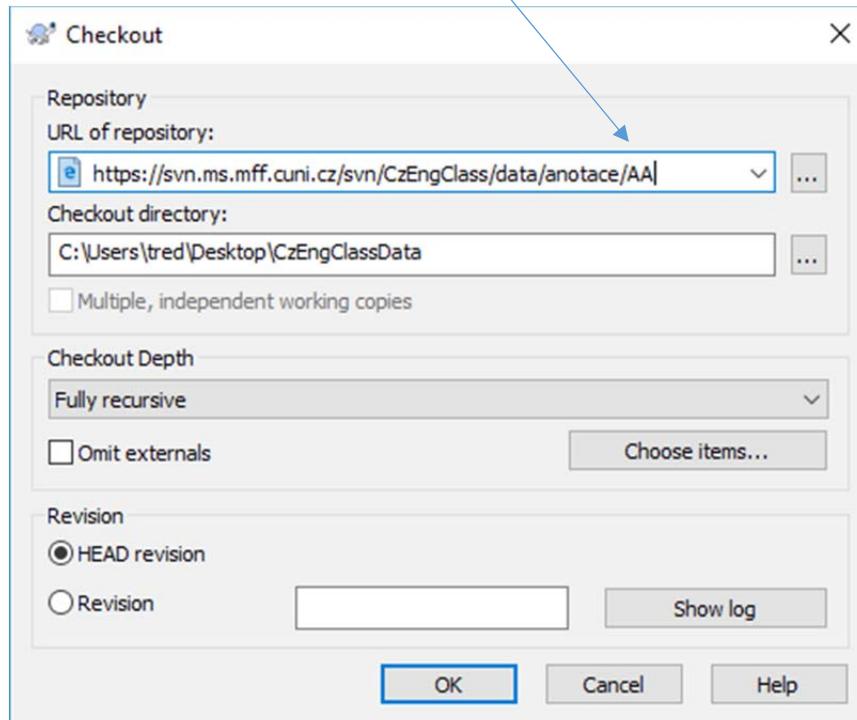
In a similar way, in the next step, you go to the ***EditorCzEngClass/resources/examples_sentences*** and extract the ***examples.zip*** file here. This way you get plenty of files labeled "***Vtext_....php***".

12.2. Data

For the annotation, each annotator needs to have its own annotation directory to store the processed data in. Therefore, at the same place as you created the ***EditorCzEngClass*** directory, you create another directory called ***CzEngClassData*** and download a copy of our annotation directory to it – use right click on this directory, select ***SVN Checkout*** and in the ***Checkout*** window into the ***URL of repository*** you fill the following addressee

<https://svn.ms.mff.cuni.cz/svn/CzEngClass/data/anotace/AA>

Instead of **AA** you fill in your own initials.



(again, replace tred in ...\\Users\\tred\\... above by your Windows username)

12.3. Configuration File Settings

In the **CzEngClassEditor\Config** directory, you create a **config_file** (copy **config_file_example**) and set the paths to the annotation file, the source files for Editor, and TrEd.

The proper path settings (Do remember to change **user_name** to your Windows username!) is as follows (DataCzEngClass can be changed to whatever place you have put your data):

ResourcePath="c:\\Users**user_name**\\Desktop\\DataCzEngClass,c:\\Users**user_name**\\Desktop\\EditorCzEngClass\\resources"

TrEdPath="c:\\Users**user_name**\\tred\\tred.bat"

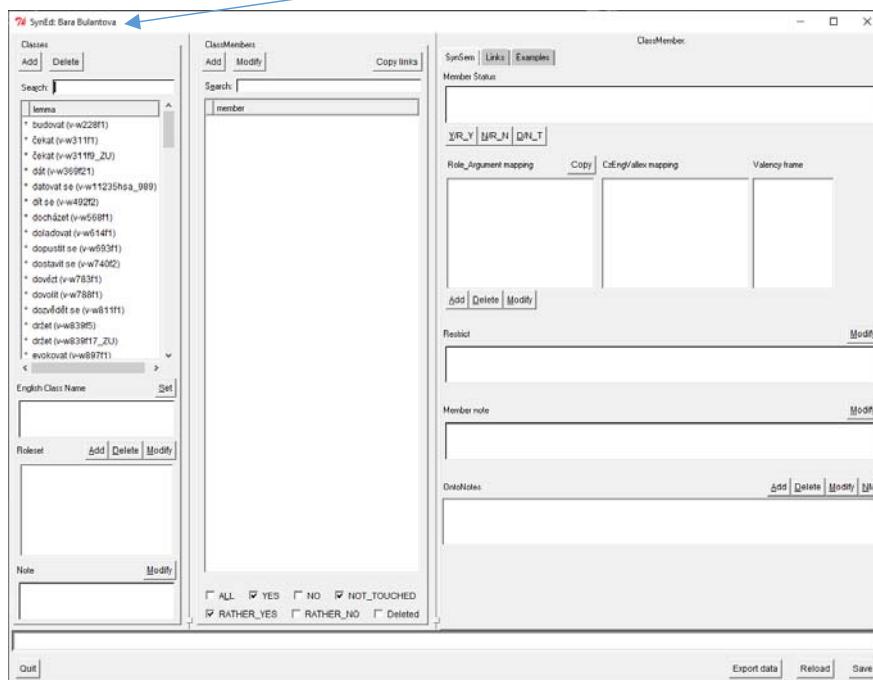
```
config_file.txt
ResourcePath="c:\\Users\\user_name\\Desktop\\DataCzEngClass,c:\\Users\\user_name\\CzEngClassEditor\\resources"
TrEdPath="c:\\Users\\user_name\\tred\\tred.bat"

;; Options changed by CzEngClass Editor on every close (DO NOT EDIT)
Geometry = 1500x1500
```

12.4. How to run CzEngClass Editor

You start to run the *CzEngClass Editor* (called *Synonyms Editor – SynEd*) by clicking the **czengclass.bat** file in the **EditorCzEngClass\CzEngClass** directory.

After opening the *SynEd* you get the following window, with anotator's name (here: Bara Bulantova) in the upper left corner.



When problems occur, you do the same as in step 1.4 - you start ***Command Prompt***, move to the ***EditorCzEngClass\czengclass*** directory, where you run ***czengclass.bat*** file. The error message should be sent to Eva Fučíková again.

SVN Update and SVN Commit

To execute ***SVN Update/SVN Commit*** commands you click the right mouse on the appropriate directory (e.g., *CzEngClass Editor*, or *DataCzEngClass*) and after the window opens, you click ***SVN Update/SVN Commit***. After the update/commit is complete, you click the ***OK*** button in the bottom pane of the *update/commit* window.

ÚFAL

ÚFAL (Ústav formální a aplikované lingvistiky; <http://ufal.mff.cuni.cz>) is the Institute of Formal and Applied linguistics, at the Faculty of Mathematics and Physics of Charles University, Prague, Czech Republic. The Institute was established in 1990 after the political changes as a continuation of the research work and teaching carried out by the former Laboratory of Algebraic Linguistics since the early 60s at the Faculty of Philosophy and later the Faculty of Mathematics and Physics. Together with the “sister” Institute of Theoretical and Computational Linguistics (Faculty of Arts) we aim at the development of teaching programs and research in the domain of theoretical and computational linguistics at the respective Faculties, collaborating closely with other departments such as the Institute of the Czech National Corpus at the Faculty of Philosophy and the Department of Computer Science at the Faculty of Mathematics and Physics.

CKL

As of 1 June 2000 the Center for Computational Linguistics (Centrum komputační lingvistiky; <http://ckl.mff.cuni.cz>) was established as one of the centers of excellence within the governmental program for support of research in the Czech Republic. The center is attached to the Faculty of Mathematics and Physics of Charles University in Prague.

TECHNICAL REPORTS

The ÚFAL/CKL technical report series has been established with the aim of disseminate topical results of research currently pursued by members, cooperators, or visitors of the Institute. The technical reports published in this Series are results of the research carried out in the research projects supported by the Grant Agency of the Czech Republic, GAČR 405/96/K214 (“Komplexní program”), GAČR 405/96/0198 (Treebank project), grant of the Ministry of Education of the Czech Republic VS 96151, and project of the Ministry of Education of the Czech Republic LN00A063 (Center for Computational Linguistics). Since November 1996, the following reports have been published.

ÚFAL TR-1996-01 Eva Hajíčová, *The Past and Present of Computational Linguistics at Charles University*
Jan Hajíč and Barbora Hladká, *Probabilistic and Rule-Based Tagging of an Inflective Language – A Comparison*

ÚFAL TR-1997-02 Vladislav Kuboň, Tomáš Holan and Martin Plátek, *A Grammar-Checker for Czech*

ÚFAL TR-1997-03 Alla Bémová at al., *Anotace na analytické rovině, Návod pro anotátory (in Czech)*

ÚFAL TR-1997-04 Jan Hajíč and Barbora Hladká, *Tagging Inflective Languages: Prediction of Morphological Categories for a Rich, Structural Tagset*

ÚFAL TR-1998-05 Geert-Jan M. Kruijff, *Basic Dependency-Based Logical Grammar*

ÚFAL TR-1999-06 Vladislav Kuboň, *A Robust Parser for Czech*

ÚFAL TR-1999-07 Eva Hajíčová, Jarmila Panevová and Petr Sgall, *Manuál pro tektogramatické značkování (in Czech)*

ÚFAL TR-2000-08 Tomáš Holan, Vladislav Kuboň, Karel Oliva, Martin Plátek, *On Complexity of Word Order*

ÚFAL/CKL TR-2000-09 Eva Hajíčová, Jarmila Panevová and Petr Sgall, *A Manual for Tectogrammatical Tagging of the Prague Dependency Treebank*

ÚFAL/CKL TR-2001-10 Zdeněk Žabokrtský, *Automatic Functor Assignment in the Prague Dependency Treebank*

ÚFAL/CKL TR-2001-11 Markéta Straňáková, *Homonymie předložkových skupin v češtině a možnost jejich automatického zpracování*

ÚFAL/CKL TR-2001-12 Eva Hajíčová, Jarmila Panevová and Petr Sgall, *Manuál pro tektogramatické značkování (III. verze)*

ÚFAL/CKL TR-2002-13 Pavel Pecina and Martin Holub, *Sémanticky signifikantní kolokace*

ÚFAL/CKL TR-2002-14 Jiří Hana, Hana Hanová, *Manual for Morphological Annotation*

ÚFAL/CKL TR-2002-15 Markéta Lopatková, Zdeněk Žabokrtský, Karolína Skwarská and Vendula Benešová,
Tektogramaticky anotovaný valenční slovník českých sloves

ÚFAL/CKL TR-2002-16 Radu Gramatovici and Martin Plátek, *D-trivial Dependency Grammars with Global Word-Order Restrictions*

ÚFAL/CKL TR-2003-17 Pavel Květoň, *Language for Grammatical Rules*

ÚFAL/CKL TR-2003-18 Markéta Lopatková, Zdeněk Žabokrtský, Karolina Skwarska, Václava Benešová, *Valency Lexicon of Czech Verbs VALLEX 1.0*

ÚFAL/CKL TR-2003-19 Lucie Kučová, Veronika Kolářová, Zdeněk Žabokrtský, Petr Pajas, Oliver Čulo, *Anotování koreference v Pražském závislostním korpusu*

ÚFAL/CKL TR-2003-20 Kateřina Veselá, Jiří Havelka, *Anotování aktuálního členění věty v Pražském závislostním korpusu*

ÚFAL/CKL TR-2004-21 Silvie Cinková, *Manuál pro tektogramatickou anotaci angličtiny*

ÚFAL/CKL TR-2004-22 Daniel Zeman, *Neprojektivity v Pražském závislostním korpusu (PDT)*

ÚFAL/CKL TR-2004-23 Jan Hajič a kol., *Anotace na analytické rovině, návod pro anotátory*

ÚFAL/CKL TR-2004-24 Jan Hajič, Zdeňka Urešová, Alevtina Bémová, Marie Kaplanová, *Anotace na tektogramatické rovině (úroveň 3)*

ÚFAL/CKL TR-2004-25 Jan Hajič, Zdeňka Urešová, Alevtina Bémová, Marie Kaplanová, *The Prague Dependency Treebank, Annotation on tectogrammatical level*

ÚFAL/CKL TR-2005-27 Jiří Hana, Daniel Zeman, *Manual for Morphological Annotation (Revision for PDT 2.0)*

ÚFAL/CKL TR-2005-28 Marie Mikulová a kol., *Pražský závislostní korpus (The Prague Dependency Treebank) Anotace na tektogramatické rovině (úroveň 3)*

ÚFAL/CKL TR-2005-29 Petr Pajas, Jan Štěpánek, *A Generic XML-Based Format for Structured Linguistic Annotation and Its application to the Prague Dependency Treebank 2.0*

ÚFAL/CKL TR-2006-30 Marie Mikulová, Alevtina Bémová, Jan Hajič, Eva Hajičová, Jiří Havelka, Veronika Kolařová, Lucie Kučová, Markéta Lopatková, Petr Pajas, Jarmila Paněnová, Magda Razímová, Petr Sgall, Jan Štěpánek, Zdeňka Urešová, Kateřina Veselá, Zdeněk Žabokrtský, *Annotation on the tectogrammatical level in the Prague Dependency Treebank (Annotation manual)*

ÚFAL/CKL TR-2006-31 Marie Mikulová, Alevtina Bémová, Jan Hajič, Eva Hajičová, Jiří Havelka, Veronika Kolařová, Lucie Kučová, Markéta Lopatková, Petr Pajas, Jarmila Paněnová, Petr Sgall, Magda Ševčíková, Jan Štěpánek, Zdeňka Urešová, Kateřina Veselá, Zdeněk Žabokrtský, *Anotace na tektogramatické rovině Pražského závislostního korpusu (Referenční příručka)*

ÚFAL/CKL TR-2006-32 Marie Mikulová, Alevtina Bémová, Jan Hajič, Eva Hajičová, Jiří Havelka, Veronika Kolařová, Lucie Kučová, Markéta Lopatková, Petr Pajas, Jarmila Paněnová, Petr Sgall, Magda Ševčíková, Jan Štěpánek, Zdeňka Urešová, Kateřina Veselá, Zdeněk Žabokrtský, *Annotation on the tectogrammatical level in the Prague Dependency Treebank (Reference book)*

ÚFAL/CKL TR-2006-33 Jan Hajič, Marie Mikulová, Martina Otradovcová, Petr Pajas, Petr Podveský, Zdeňka Urešová, *Pražský závislostní korpus mluvené češtiny. Rekonstrukce standardizovaného textu z mluvené řeči*

ÚFAL/CKL TR-2006-34 Markéta Lopatková, Zdeněk Žabokrtský, Václava Benešová (in cooperation with Karolína Skwarska, Klára Hrstková, Michaela Nová, Eduard Bejček, Miroslav Tichý) *Valency Lexicon of Czech Verbs. VALLEX 2.0*

ÚFAL/CKL TR-2006-35 Silvie Cinková, Jan Hajič, Marie Mikulová, Lucie Mladová, Anja Nedolužko, Petr Pajas, Jarmila Paněnová, Jiří Semecák, Jana Šindlerová, Josef Toman, Zdeňka Urešová, Zdeněk Žabokrtský, *Annotation of English on the tectogrammatical level*

ÚFAL/CKL TR-2007-36 Magda Ševčíková, Zdeněk Žabokrtský, Oldřich Krůza, *Zpracování pojmenovaných entit v českých textech*

ÚFAL/CKL TR-2008-37 Silvie Cinková, Marie Mikulová, *Spontaneous speech reconstruction for the syntactic and semantic analysis of the NAP corpus*

ÚFAL/CKL TR-2008-38 Marie Mikulová, *Rekonstrukce standardizovaného textu z mluvené řeči v Pražském závislostním korpusu mluvené češtiny. Manuál pro anotátory*

ÚFAL/CKL TR-2008-39 Zdeněk Žabokrtský, Ondřej Bojar, *TectoMT, Developer's Guide*

ÚFAL/CKL TR-2008-40 Lucie Mladová, *Diskurzní vztahy v češtině a jejich zachycení v Pražském závislostním korpusu 2.0*

ÚFAL/CKL TR-2009-41 Marie Mikulová, *Pokyny k překladu určené překladatelům, revizorům a korektorům textů z Wall Street Journal pro projekt PCEDT*

ÚFAL/CKL TR-2011-42 Loganathan Ramasamy, Zdeněk Žabokrtský, *Tamil Dependency Treebank (TamilTB) – 0.1 Annotation Manual*

ÚFAL/CKL TR-2011-43 Nguy Giang Linh, Michal Novák, Anna Nedoluzhko, *Coreference Resolution in the Prague Dependency Treebank*

ÚFAL/CKL TR-2011-44 Anna Nedoluzhko, Jiří Mírovský, *Annotating Extended Textual Coreference and Bridging Relations in the Prague Dependency Treebank*

ÚFAL/CKL TR-2011-45 David Mareček, Zdeněk Žabokrtský, *Unsupervised Dependency Parsing*

ÚFAL/CKL TR-2011-46 Martin Majliš, Zdeněk Žabokrtský, *W2C – Large Multilingual Corpus*

ÚFAL TR-2012-47 Lucie Poláková, Pavlína Jínová, Šárka Zikánová, Zuzanna Bedřichová, Jiří Mírovský, Magdaléna Rysová, Jana Zdeňková, Veronika Pavlíková, Eva Hajičová, *Manual for annotation of discourse relations in the Prague Dependency Treebank*

ÚFAL TR-2012-48 Nathan Green, Zdeněk Žabokrtský, *Ensemble Parsing and its Effect on Machine Translation*

ÚFAL TR-2013-49 David Mareček, Martin Popel, Loganathan Ramasamy, Jan Štěpánek, Daniel Zemana, Zdeněk Žabokrtský, Jan Hajič *Cross-language Study on Influence of Coordination Style on Dependency Parsing Performance*

ÚFAL TR-2013-50 Jan Berka, Ondřej Bojar, Mark Fishel, Maja Popović, Daniel Zeman, *Tools for Machine Translation Quality Inspection*

ÚFAL TR-2013-51 Marie Mikulová, *Anotace na tektogramatické rovině. Dodatky k anotátorské příručce (s ohledem na anotování PDTSC a PCEDT)*

ÚFAL TR-2013-52 Marie Mikulová, *Annotation on the tectogrammatical level. Additions to annotation manual (with respect to PDTSC and PCEDT)*

ÚFAL TR-2013-53 Marie Mikulová, Eduard Bejček, Jiří Mírovský, Anna Nedoluzhko, Jarmila Panevová, Lucie Poláková, Pavel Straňák, Magda Ševčíková, Zdeněk Žabokrtský, *Úpravy a doplňky Pražského závislostního korpusu (Od PDT 2.0 k PDT 3.0)*

ÚFAL TR-2013-54 Marie Mikulová, Eduard Bejček, Jiří Mírovský, Anna Nedoluzhko, Jarmila Panevová, Lucie Poláková, Pavel Straňák, Magda Ševčíková, Zdeněk Žabokrtský, *From PDT 2.0 to PDT 3.0 (Modifications and Complements)*

ÚFAL TR-2014-55 Rudolf Rosa, *Depfix Manual*

ÚFAL TR-2014-56 Veronika Kolářová, *Valence vybraných typů deverbalivních substantiv ve valenčním slovníku PDT-Vallex*

ÚFAL TR-2014-57 Anna Nedoluzhko, Eva Fučíková, Jiří Mírovský, Jiří Pergler, Lenka Šíková, *Annotation of coreference in Prague Czech-English Dependency Treebank*

ÚFAL TR-2015-58 Zdeňka Urešová, Eva Fučíková, Jana Šindlerová, *CzEngVallex: Mapping Valency between Languages*

ÚFAL TR-2015-59 Kateřina Rysová, Magdaléna Rysová, Eva Hajičová, *Topic–Focus Articulation in English Texts on the Basis of Functional Generative Description*

ÚFAL TR-2016-60 Kira Droganova, Daniel Zeman, *Conversion of SynTagRus (the Russian dependency treebank) to Universal Dependencies*

ÚFAL TR-2018-61 Lukáš Kyjánek, *Morphological Resources of Derivational Word-Formation Relations*

ÚFAL TR-2019-62 Zdeňka Urešová, Eva Fučíková, Eva Hajičová, *CzEngClass: Contextually-based Synonymy and Valency of Verbs in a Bilingual Setting (CzEngClass: Kontextová synonymie a valence sloves v bilingvním prostředí)*