

MATEMATICKO-FYZIKÁLNÍ FAKULTA
PRAHA

**Annotation guidelines for German verbal synonyms included in
SynSemClass Lexicon**

ZDEŇKA UREŠOVÁ, EVA FUČÍKOVÁ, JAN HAJIČ, KAROLINA ŻACZYNSKA

ÚFAL Technical Report
TR-2021-70

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 German verbal synonyms included in SynSemClass Lexicon

Zdeňka Urešová, Eva Fučíková, Jan Hajič, Karolina Zaczynska

uresova,fucikova,hajic@ufal.mff.cuni.cz, Karolina.Zaczynska@dfki.de

Acknowledgements

This work described herein is supported by the grant LUSyD GX20-16819X of the Czech Science Foundation, and it uses resources hosted by the LINDAT/CLARIAH-CZ Research Infrastructure, project No. LM2018101, supported by the Ministry of Education of the Czech Republic. The work on German is supported in part by the grant HumanE AI Network supported by the EC by award No. 952026.

Abstract

This manual is created as part of a project called Multilingual Event-type-anchored Ontology for Natural Language Understanding (META-O-NLU) which aims to extend an already existing bilingual Czech-English synonym lexicon (SynSemClass) with a new language, i. e., German, and make the resource multilingual. This will enable comparison of semantic and syntactic properties across languages and provide data for NLP experiments.

The SynSemClass lexicon (available online¹) groups Czech and English contextually-based synonymous verbs by semantic and syntactic properties into bilingual synonym classes. Each synonymous class is characterized both meaning-wise (semantic roles) and structurally (valency arguments) by linking (mapping) semantic roles and valency members (Role_Argument mapping). Each class is assigned a set of semantic roles (Roleset) and the prototypical meaning of the synonym class is represented by the English and Czech verb sense. The individual entries (Class Members - CMs) are further linked to other lexical resources, like PDT-style valency lexicons, WordNet, FrameNet, VerbNet and PropBank, OntoNotes, and exemplified by the real translational texts taken from a parallel corpus. For Czech and English synonyms the Prague Czech-English Dependency Treebank (PCEDT)² is used.

This manual follows a manual for annotation of English and Czech synonyms³ and suggests how to annotate German verbs.

¹ <https://lindat.mff.cuni.cz/services/SynSemClass/>

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

³ Urešová Zdeňka, Fučíková Eva, Hajičová Eva: *CzEngClass: Contextually-based Synonymy and Valency of Verbs in a Bilingual Setting*. Technical report no. 2019/62, Copyright © ÚFAL MFF UK, Prague, Czechia, ISSN 1214-5521, 101 pp., 2019

Contents

1	Goal of the Annotation	3
2	Annotation Tasks	4
3	Class Member Status annotation	5
4	Role_Argument mapping	6
5	Links to Other Lexical Sources	9
5.1	Links to Woxikon	11
5.2	Links to VALBU and GUP	12
5.2.1	E-VALBU-lexicon	12
5.2.2	GUP – German Universal Proposition Bank	14
5.3	Links to FrameNet des Deutschen	17
6	Examples	22

1 Goal of the Annotation

The aim of the annotation is to enrich a lexicon of bilingual Czech-English synonyms with German synonymous verbs. German synonyms are extracted from the sentence-aligned parallel English-German corpus ParaCrawl⁴. A functionally adequate relationship in terms of translation must exist between the meaning of the English, Czech, and German verb, i.e., the English, the Czech, and the German verb must be synonymous in the given context(s). All automatically prepared German verbs of a given class must have the same set (Roleset) of semantic roles. Ex.: Absorber, Absorbed – the class *pohltit* (*v-w3678f1*) *absorb* (*ev-w14f1*)⁵. The German translational verb sense should have similar valency to the Czech and English verbs in the given sense.

For the annotation, a special annotation editor – called Synonyms Editor (SynEd) – is used. SynEd captures (**Fig. 1**, from left to right) Classes, Class Members and syntactic-semantic information (SynSem), links (Links) and examples (Examples). The SynEd tries to make the work of the annotator as simple as possible.

⁴ <https://paracrawl.eu/>

⁵ <https://lindat.mff.cuni.cz/services/SynSemClass/SynSemClass.html>

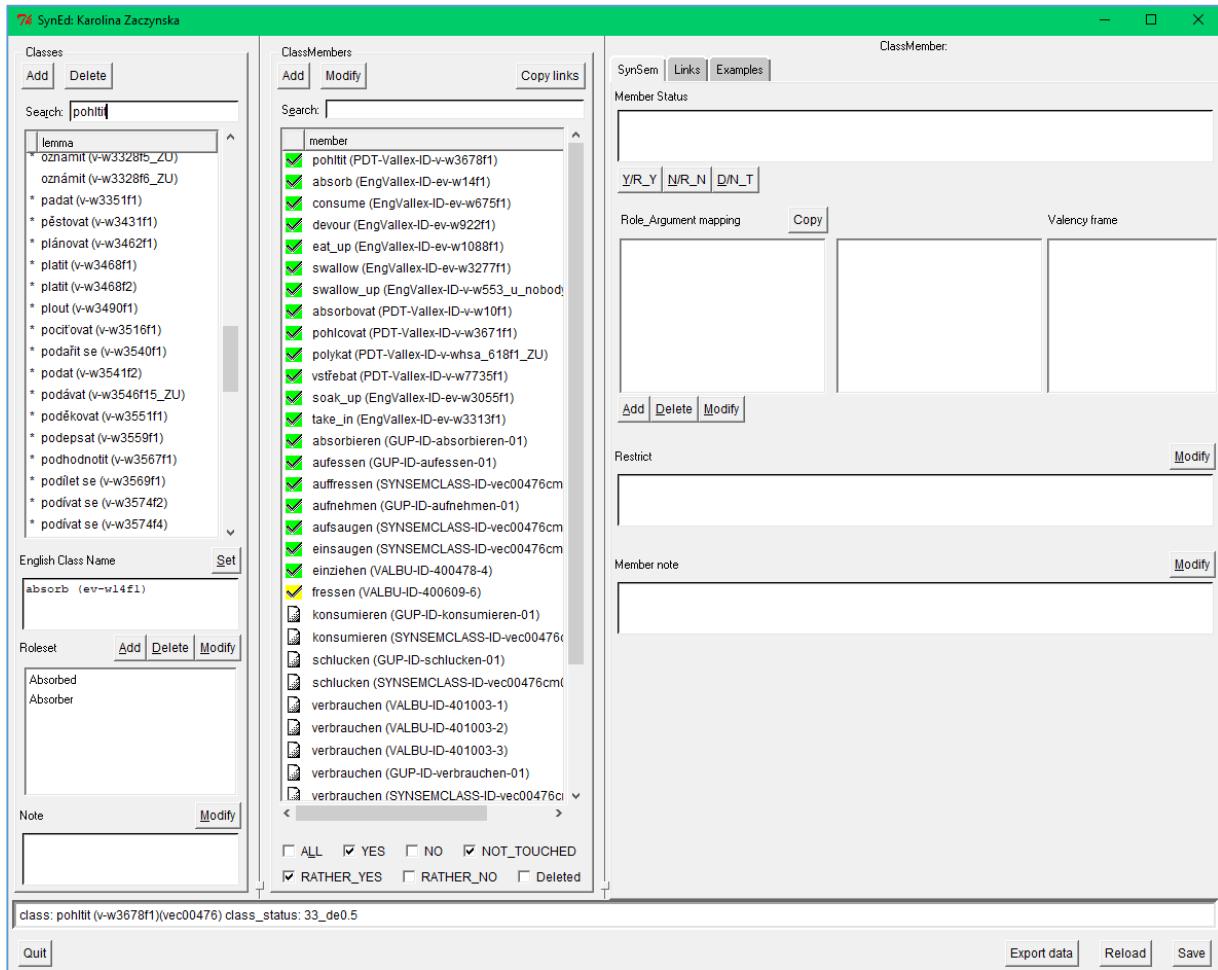


Fig. 1 SynEd – Classes, Class Members and syntactic-semantic information (SynSem), links (Links) and examples (Examples).

2 Annotation Tasks

The bilingual classes of verb synonyms based on the translation from English to Czech are already established. Besides the Cz and En Class Members (CMs), every class contains its specific "Roleset" (a set of semantic roles) and each CM in one synonym class is described by the following characteristics:

- by mapping semantic roles from this Roleset to valency,
- by linking to other lexical (internal and external) resources, and
- by assigning representative sentences from a parallel corpus.

As the annotation of German verbs builds on already existing CZ-EN annotations, the Roleset for each synonym class is already defined, so it does not need to be annotated. But the annotator

has to go manually through the automatically generated German verbal CMs candidates and for every German CM candidate annotate the following tasks:

- the Class Member Status
- the Role_Argument mapping (mapping of semantic roles from the given Roleset to valency frame of each German verb)
- Links to external resources (Woxikon, VALBU, GUP, FN des Deutschen)
- Examples from Paracrawl corpus.

The annotator has to take into account that a single verb can have multiple meanings. Those are – in the context of the PCEDT – determined by a valency frame. While valency frames are already specified for the Czech and English verbs in the text and captured in the associated valency lexicons (CzEngVallex⁶, PDT-Vallex⁷, and EngVallex⁸), valency frames for German verbs are not so straightforward, therefore we describe the rules for expressing valency for German in more detail below.

3 Class Member Status annotation

CM Status is used to mark whether or not the verb sense is included in the class. For this annotation, you select one of the following values in the Tab “SynSem” and fill it into Member Status Tab.

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

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

If unsure, you 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 sense 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.

Restriction box in the editor is used for inserting a condition that a given verb must fulfill if it is a member of a given synonymous class. Here, only free text explaining in what context the German verb can be considered synonymous with EN and CZ verbs is inserted. There are no

⁶ <https://lindat.mff.cuni.cz/services/CzEngVallex/>

⁷ <http://lindat.mff.cuni.cz/services/PDT-Vallex/>

⁸ <http://lindat.mff.cuni.cz/services/EngVallex/>

format requirements. One Example would be for the class vec00347 “accuse” the CM lemma “geben” where the restriction is “Schuld” for the phrase “jmd. die Schuld geben”.

Notes box in the editor is used for notes to the individual CMs. Here the annotator can add notes and info like a negative/positive connotation of the CM in contrast to the other entries for that class, or that the CM is colloquial, etc.

4 Role_Argument mapping

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. This means that for each CM of a synonymous class X, the semantic roles contained in Roleset need to be meaningfully mapped to valency members (Arguments and obligatory Adjuncts) of the valency frame of this CM.

When mapping the roles from a given Roleset, each role must be mapped to “SOMETHING” from the valency frame of a verb in that class. The “SOMETHING” could be

- either a member of the valency frame
- or a proxy functor (#any⁹, #sb, #sth, #smh, #smt, #swh)

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.

In cases, where we are able to use a functor (for German verbs it means A0, A1... or V0, V1...) because we believe that those should be included in the appropriate German valency frame for the given verb (if this formal valency frame exists), we prefer to use the functor. But in cases, where we believe that the "meaning slots" would not be included as valency slots in valency frame, we will use a proxy functor (#any, #sb, #sth, #smh, #smt, #swh).

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.

In SynEd, we annotate this mapping under the Syntactic-Semantic Information tab (SynSem) in the Role_Argument mapping window (**Fig. 2**).

⁹ “#any” means “sb” or “sth” (e.g., “Jane/The winter has come.”), not “any functor

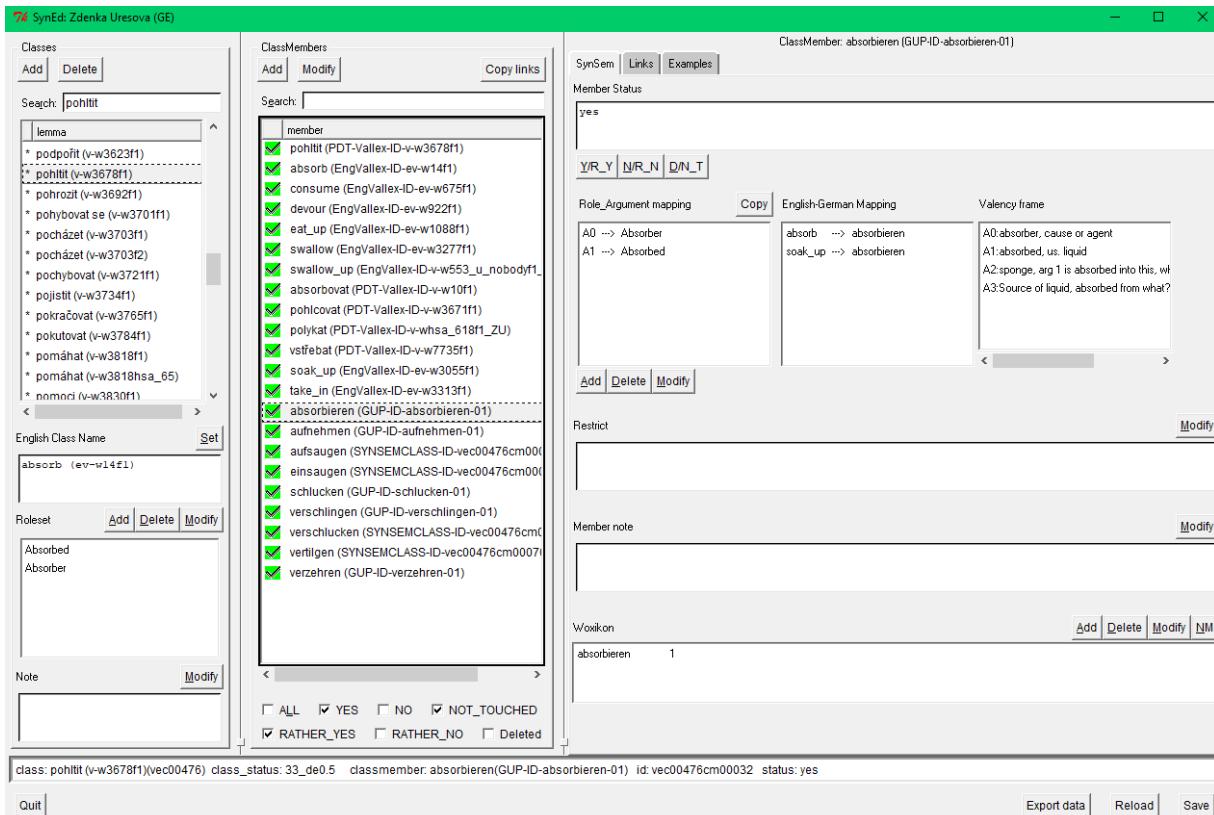


Fig. 2 SynSem tab with windows: Role_Argument mapping, English-German Mapping and Valency frame for CM *absorbieren* in the *pohltit/absorb* class.

First, we open the SynSem tab (**Fig. 2**) in SynEd, where the following information is already displayed about the valency frame (VF) for each class member (*absorbieren*) of the synonymous class (*pohltit/absorb*) – Valency frame window (A0, A1, A2, A3) and the mapping of this CM to its English translational counterparts displayed in English-German mapping window (*absorbieren – absorb, absorbieren – soak up*).

The English-German 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. 3**).

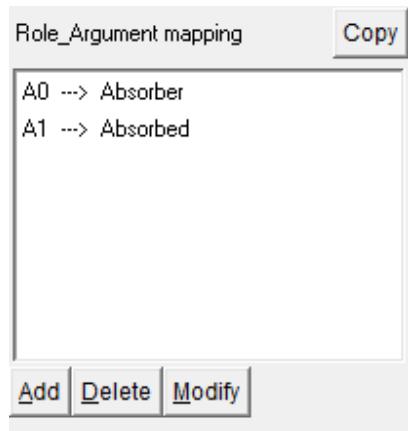


Fig. 3 The Role_Argument mapping window (CM *absorbieren* class *pohltit/absorb*) – edit buttons lower left and copy mapping button upper right

For this annotation, use Add to open the mapping window (**Fig. 4**) and gradually map the valency members of the CM's of the synonymous class to the prepared roles via the OK+Next button.

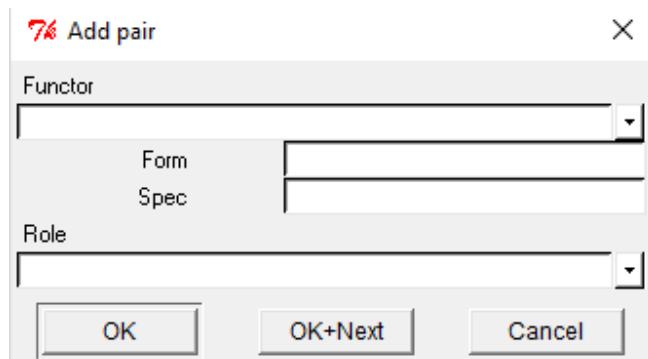


Fig. 4 The editing window for mapping

If you want to restrict a mapped valency member (*A3*) to a specific form (*A3 restricted to the preposition von*), fill that form in the field Form, as shown in **Fig 4a**.

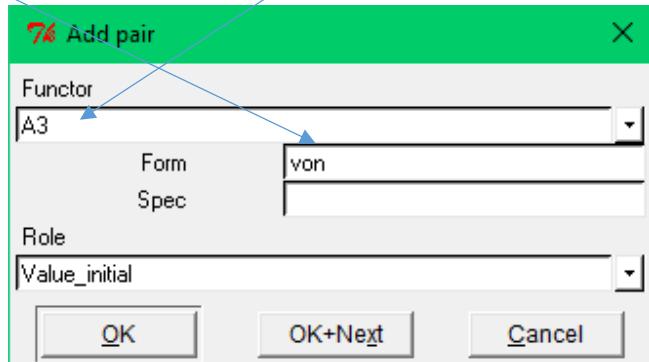


Fig.4a The valency member form annotation

The **Fig. 4b** depicts the Role_argument mapping window with a final annotation of the restricted forms *um*, *von*, and *zu* for *A2*, *A3*, and *A4* valency members.

Role_Argument mapping		Copy
A0	---> Agent	
A1	---> Item_changed	
A3(von)	---> Value_initial	
A4(zu)	---> Value_final	
A2(um)	---> Difference	

Fig. 4b The annotation of forms for *A2*, *A3*, and *A4* valency members

5 Links to Other Lexical Sources

We want references to other lexical sources to be part of the SynSemClass lexicon 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.

The aim of the annotation of German CMs is to assign the following links in SynEd, recorded in the Links tab, to individual Class Members (CMs) (**Fig. 5 – Links**):

Woxikon, VALBU, GUP, FN des Deutschen.

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

The screenshot shows the SynEd application interface with the title bar "SynEd: Zdenka Uresova (GE)". The main window is divided into several panes:

- Left pane (Classes):** Shows a search bar for "pohltit" and a list of lemmas including "pohltit", "absorb", "consume", "devour", "eat_up", "swallow", "swallow_up", "absorbat", "pohlcovat", "polykat", "vstfebat", "soak_up", "take_in", and "absorbieren".
- Middle-left pane (ClassMembers):** Shows a search bar for "member" and a list of member links for "pohltit" and "absorbieren".
- Right pane (FrameNet Des Deutschen):** Shows a table for "aufsaugen" with count 511. It includes tabs for "SynSem", "Links" (selected), and "Examples".
- Bottom pane (ParaCrawl German):** Shows a table comparing "absorbieren" and "absorb" across different datasets.

At the bottom, there is a status bar with the text "class: pohltit (v-w3678f1)(vec00476) class_status: 33_de0.5 classmember: absorbieren(GUP-ID-absorbieren-01) id: vec00476cm00032 status: yes" and buttons for "Export data", "Reload", and "Save".

Fig. 5 Showing links in the SynEd editor for the verb *absorbieren* in the *pohltit/absorb* class

5.1 Links to Woxikon

We open the Woxikon website <https://synonyme.woxikon.de/> by browsing for the appropriate German CM and fill in the appropriate sense Link into the **Woxikon** window which is placed under the **SynSem Tab** (**Fig. 6a**).

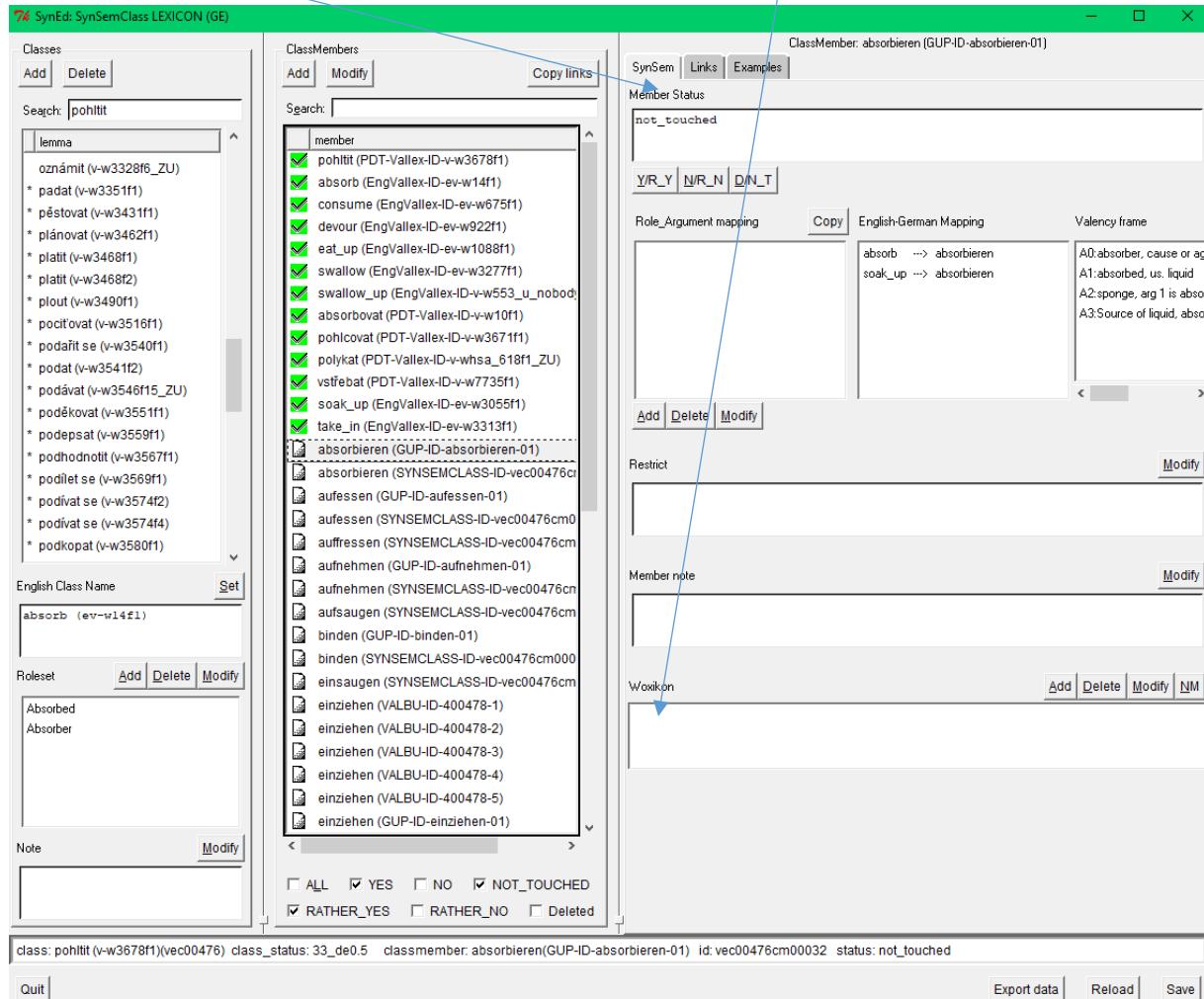


Fig. 6a SynEd with “opened” **SynSem** tab where the **Woxikon** window for links is placed

The correct Link (**Sense**) to Woxikon is based on the Roleset, CZ and En CMs and has to be filed by opening the **Add Woxikon window** for the appropriate CM **Lemma** (**Fig. 6b**).

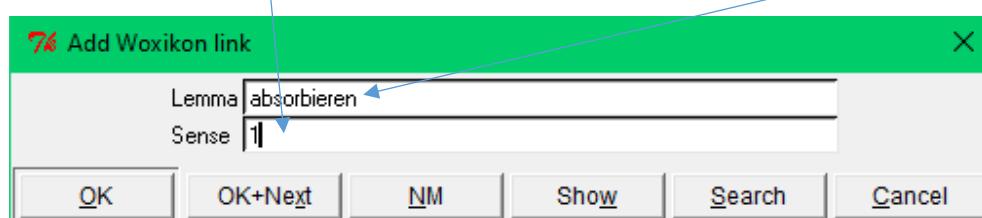


Fig. 6 Selected sense from Woxikon for the verb *absorbieren*

5.2 Links to VALBU and GUP

Since there are no online available German valency lexicons distinguishing the individual verb senses by a formal valency frame, there exist more CM entries for one verb lemma in each class, in order we can include different senses. One entry = one valency frame = one sense. For Cz and En verbs we differentiate the senses by valency frames. But since we do not have the appropriate valency frames for German, we decided to include Class Member candidates with three kinds of IDs representing the individual senses in the SynEd, where for 1 and 2 the external links are already given by the editor:

- Class Member with VALBU-ID
- Class Member with GUP-ID
- Class Member with SYNSEMCLAS-ID

We use VALBU and GUP as auxiliary resources (see below).

For example, for the "*einziehen*" lemma in *pohltit/absorb*(vec00476) class, there are 7 entries: 4 with VALBU-ID, one with GUP-ID, and one extra with SYNSEMCLAS-ID.

Because some of the included German entries/senses may overlap, we suggest the following preference for equivalent senses, the status "YES" is given to the VALBU-ID member, if it is not available, the member with the GUP-ID comes up, and then the SynSemClass-ID Class Member.

Given there are synonymous candidates with both VALBU-ID and GUP-ID, the VALBU-ID has to be preferred and gets the Status YES. The GUP-ID CM gets the "No" status and into the Member Note tab the annotator writes "same as VALBU-ID-...". When there is no entry with VALBU or GUP-ID, the SYNSEMCLASS-ID has to be used.

For German, external resources we annotate links to are

- **E-VALBU-lexicon** – Elektronisches Valenzwörterbuch deutscher Verben¹⁰ and to
- **GUP-database** – German Universal Proposition Bank¹¹.

We adapted both resources to suit our needs.

5.2.1 E-VALBU-lexicon

For E-VALBU, we decided to include the individual senses (matching the individual "Lesarten" listed) bearing a E-VALBU-ID, where the number following the ID indicates the link to the appropriate verb sense in VALBU lexicon.

¹⁰ <https://grammis.ids-mannheim.de/verbvalenz>

¹¹ <http://alanakbik.github.io/multilingual.html>; https://github.com/UniversalDependencies/UD_German-GSD

For example, for the CM candidate sense “schieben(VALBU-ID-400863-2)”, the no. 400863 is the lexeme-ID and the no. 2 links this sense to the appropriate *Lesart* in VALBU: <https://grammis.ids-mannheim.de/verbs/view/400863/2>.

Next, we prepared an “artificial” formal valency frame for each sense based on the verbal exemplification (*Strukturbispiel* in E-VALBU) of the individual valency slots (*Satzbauplan*). This Valency slot was prepared by translating the *Satzbauplan* into argument-slots (for E-VALBU VA01, VA02, etc.) and can be selected by the drop-down menu in Role_Argument mapping in the editor.

For example, the CM candidate sense “*schieben(VALBU-ID-400863-2)*” based on E-VALBU (Fig. 7) contains three valency slots: V0, V1, V2, that are recorded in the SynEd in the Valency frame window (Fig. 8).

schieben auf (Lesart 2)

Strukturbispiel

jemand schiebt etwas auf jemanden/etwas

Im Sinne von

jemand legt etwas jemandem/etwas zur Last oder macht jemanden/etwas für etwas verantwortlich

Satzbauplan

K_{sub}, K_{akk}, K_{prp}

Fig. 7. E-VALBU characteristics (*Strukturbispiel* and *Satzbauplan*) for *Lesart 2* of the verb *schieben*

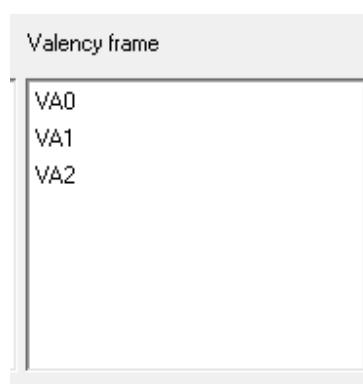


Fig.8 Formal valency frame for CM with E-VALBU-ID

5.2.2 GUP – German Universal Proposition Bank

For GUP, we decided to include the individual senses (matching the individual “Roleset id”) bearing a GUP ID, where the number following the ID indicates the link to the appropriate verb sense in the GUP database.

One example would be the CM candidate sense “*vorwerfen(GUP-ID-vorwerfen-01)*”, where the no. 01 links this sense to the appropriate Roleset in GUP
http://alanakbik.github.io/UniversalPropositions_German/vorwerfen.html#accuse.01.

Next, we included also a formal valency frame for each sense based on the Roles exemplification for each “Roleset id” in GUP. This Valency slot is already prepared-

For example, the CM candidates sense “*vorwerfen(GUP-ID-vorwerfen-01)*” contains three valency slots: **A0, A1, A2** based on GUP characteristics (**Fig. 9**). This valency frame is recorded in the SynEd in the Valency frame window (**Fig. 10**).

Predicate: *vorwerfen*

Roleset id: *vorwerfen.01*, accuse, accuse of a crime **Source:** [accuse.01](#) (*accuse.01, blame.01, reproach.01*)

vorwerfen.01: jdm. etw. vorwerfen

Roles:

A0: *accuser*
A1: *accused*
A2: *crime*

Fig. 9 GUP characteristics (Roleset id and Roles for sense 01 of the verb *vorwerfen*)



Fig. 10 Formal valency frame for CM with GUP ID

In addition to CMs with E-VALBU-IDs and GUP-IDs, you can also find members with SYNSEMCLAS-IDs among the German class member verb candidates in SynEd – Class Members with SYNSEMCLAS-IDs are uploaded to the editor for cases where the verb sense (i.e., CM) is not included in either E-VALBU or GUP.

By using Enter button staying on the individual German CM you open the equivalent source with pre-annotated links:

- for VALBU-ID CMs <https://grammis.ids-mannheim.de/>
- for GUP-ID http://alanakbik.github.io/UniversalPropositions_German/index.html

You have to check the links and assess their accuracy relative to the given synonymous class, cf., Fig. 9 for E-VALBU links and Fig. 10 for GUP links.

SynEd: Zdenka Uresova (GE)

The screenshot shows the SynEd application interface. On the left, the 'Classes' panel displays a search for 'vinit' with results including 'accuse' (evr=w31f1). The 'ClassMembers' panel shows a search for 'member' with results including 'schieben'. In the center, the 'FrameNet Des Deutschen' panel shows 'NO MAPPING' for 'schieben'. A blue arrow points from the 'schieben' entry in the 'ClassMembers' panel to the 'Edit VALBU link' dialog box. This dialog box is titled 'Edit VALBU link' and contains fields for Lemma ('schieben'), ID ('400863'), and Sense ('2'). The 'OK' button is highlighted. At the bottom of the interface, there are buttons for 'Export data', 'Reload', and 'Save'.

Fig. 9 Link to E-VALBU for the verb *schieben* in the class *vinit/accuse*

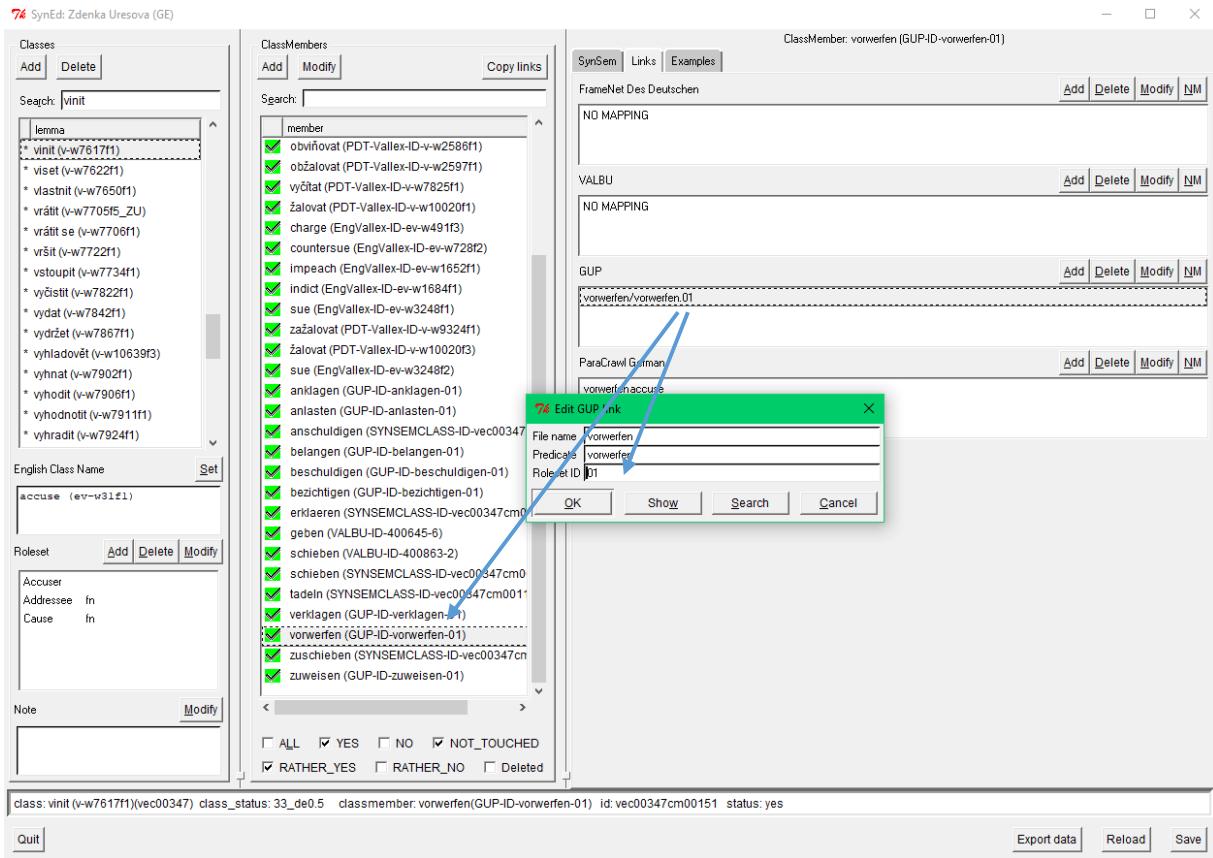


Fig. 10 Link to GUP for the verb *vorwerfen* in the class *vinit/accuse*

5.3 Links to FrameNet des Deutschen

From the point of view of semantic roles, the most important source and inspiration for SynSemClass lexicon is the FrameNet electronic database¹² which works with semantic roles (frame elements – FEs) within semantic frames.

There exist German versions of FrameNet based on the Berkeley FrameNet, namely German FrameNet¹³ and FrameNet des Deutschen¹⁴. Our lexicon makes use of **FrameNet des Deutschen (FNdD)**. However, this resource is limited and not all German verbs are included by now. Also this source is not searchable by lexical unit but by frames and frame elements only. For our purposes we must improvise a bit, therefore see steps 1, 2 and 3 bellow.

¹² <https://framenet.icsi.berkeley.edu/fndrupal/WhatIsFrameNet>

¹³ <http://www.laits.utexas.edu/gframenet/>

¹⁴ <https://gsw.phil.hhu.de/framenet/>

Step 1

First, it is helpful to check the FN links for En CMs in the given class, for example the FN link for the CM *absorb* in *pohlit/absorb* class (Fig. 11a, 11b) has a link to *Soaking up Frame*¹⁵ in the Berkeley FN.

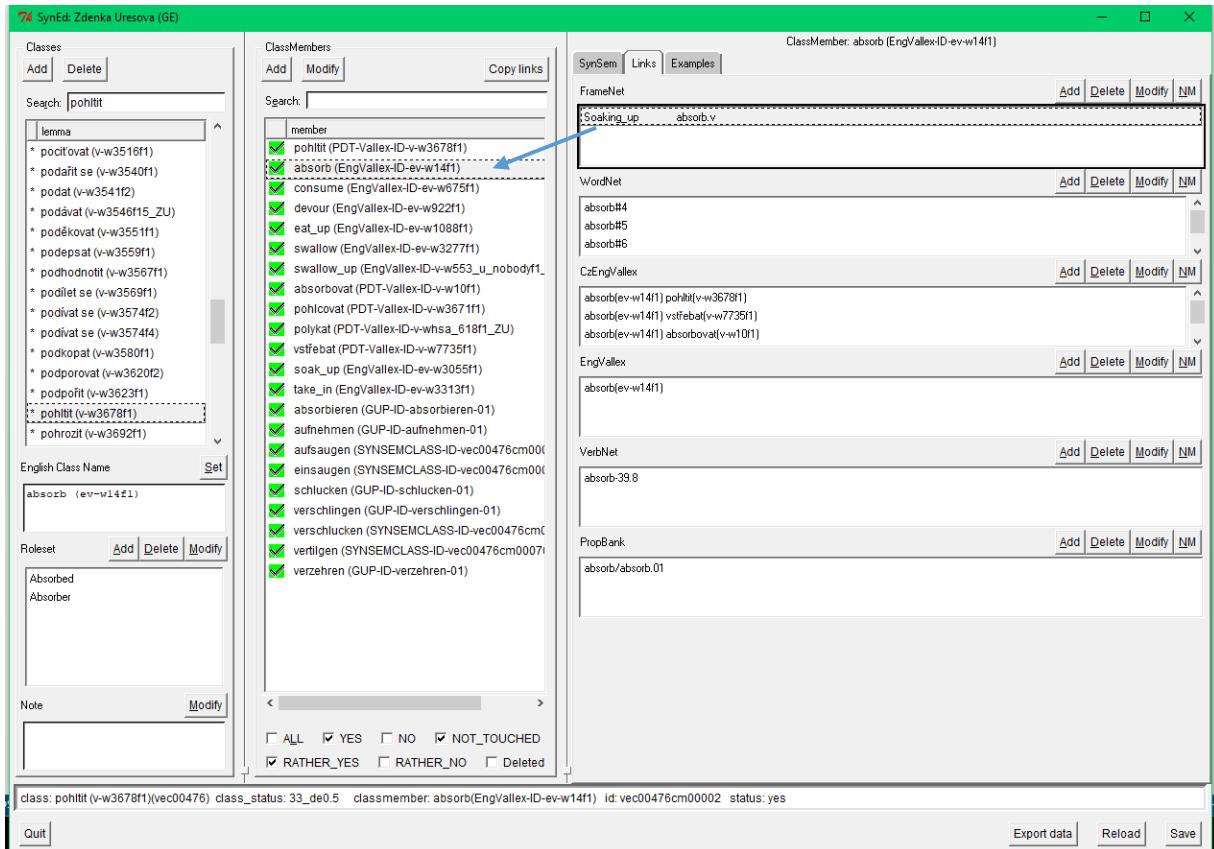


Fig. 11a Link to the Berkley FN for *absorb* in *pohlit/absorb* class to *Soaking up Frame*

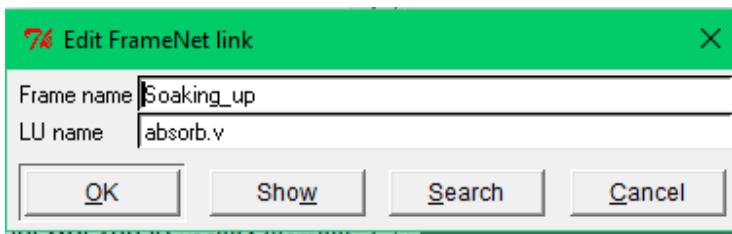


Fig. 11b Editing window for *absorb* in *pohlit/absorb* class (Fig. 11a, 11b) has a link to *Soaking up Frame*.

¹⁵ https://framenet2.icsi.berkeley.edu/fnReports/data/frameIndex.xml?frame=Soaking_up&banner=

Step 2

Next, since FNdD is searchable by Frame Index and FE Index only and not by lexical unit (LU), we must improvise and try to search for the FN link described to any En CM in the class (Fig. 12) and check if there exists any analogical Frame which would be suitable for the German CM in question.

The screenshot shows the 'Frame Index' search interface. At the top, there is a search bar with a magnifying glass icon and the word 'soak'. Below the search bar are two buttons: 'OR' and 'AND', with a toggle switch between them. To the right of the search bar is a 'Filter' button. The main results area is divided into sections by category symbols (#A and #E). Section #A contains the result 'Aufsaugen (urspr.: BFN Soaking_up)' with a 'FN-1.7' label. Section #E contains the result 'Einweichen (urspr.: BFN Soaking)' with a 'FN-1.7' label.

Fig. 12 Search in FNdD for the FN link described to *absorb* in *pohlit/absorb* class.

After checking the searched results, it is up to the annotator to choose the appropriate Frame Link to German CM (Fig. 12). For example, the FNdD link for the CM *absorbieren* in *pohlit/absorb* class (Fig. 13) has a link to the *Aufsaugen* Frame¹⁶.

¹⁶ <https://gsw.phil.hhu.de/framenet/frame?id=511>

☰ > Frame Index > Aufsaugen

Frame: Aufsaugen

Definition:

Ein Element füllt sich mit einer Substanz durch die Bewegung der Flüssigkeit in sichtbare oder unsichtbare Zwischenräume im Element.

Papier und Karton ABSORBIEREN Feuchtigkeit im Winter .

Ein stabiler Boden SAUGT Wasser auf.

Externe Ressourcen:

- Soaking_up (Berkeley FrameNet) ↗

Tags:

FN-1.7 Korrektur: Bearbeitung abgeschlossen

Fig. 13 FrameNet des Deutschen for *absorbieren* in *pohlit/absorb*

Step 3

Once the appropriate Frame Link to German CM from FNDD is determined, the annotator has to fill in this Link (or Links) into the FrameNet des Deutschen box. It is done by clicking the Add button next to FrameNet des Deutschen (Fig. 14) in the tab **Links**.

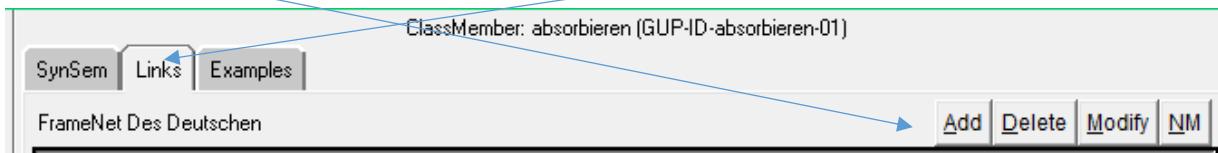


Fig. 14 Add button next to FrameNet des Deutschen

This opens the editing window for links to FNDD (Fig. 15) where you have to fill in the **Frame ID** and **Frame name** (Fig. 16) according to FN des Deutschen web page (Fig. 17).

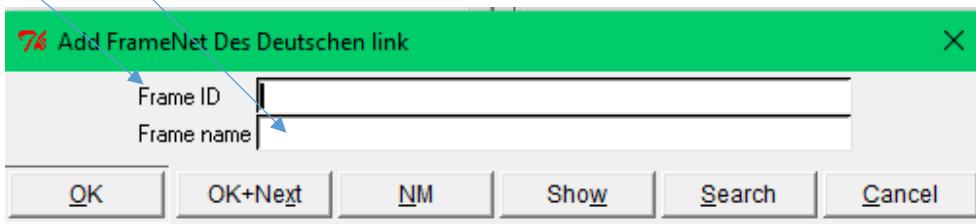


Fig. 15 The editing window for links to FNDD

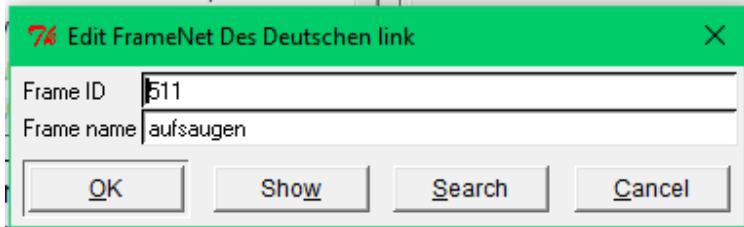


Fig. 16 The Frame ID and Frame name filled

The **Frame ID** is recorded in the **URL** and the **Frame name** is next to **Frame** (Fig. 17)

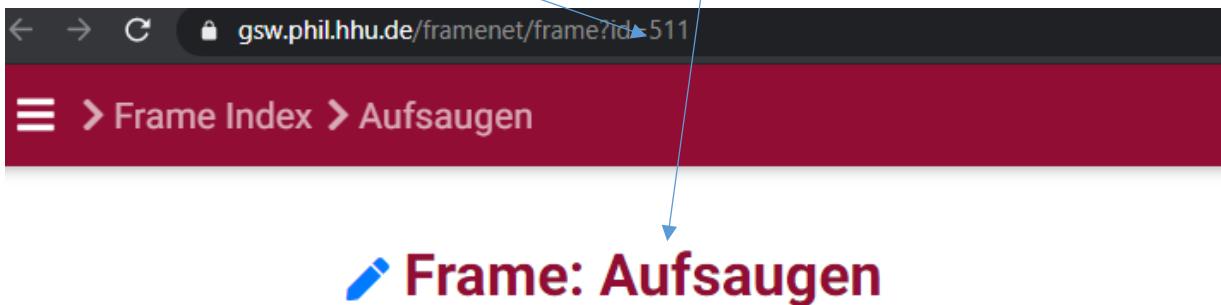


Fig. 17 The Frame ID 511 and Frame name Aufsaugen at the FN des Deutschen web page

The final annotation of a Link to FN des Deutschen for *absorbieren* of *pohltit/absorb* class is captured in Fig. 18.

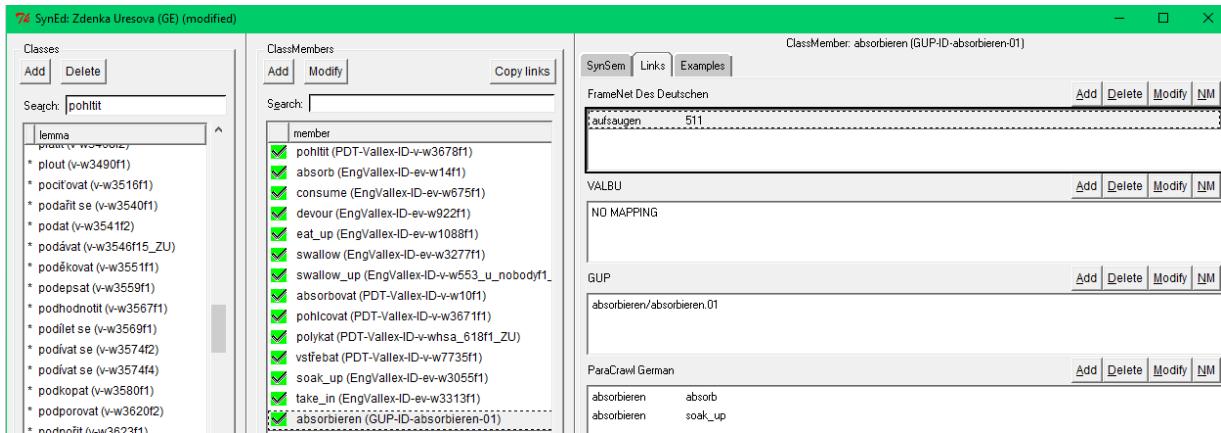


Fig. 18 Link to FN des Deutschen for *absorbieren* of *pohltit/absorb* class

In SynSemClass lexicon, we try to add links to FNND for all German class members not only for those that actually occur in the FNND frames examples. Thus, we will help our

colleagues from FndD to build a list of lexical units that might serve as a good basis for their work.

6 Examples

There are automatically inserted example sentences from the Paracrawl corpus for each German CM candidate; in the SynEd, examples are under a specific Tab called **Examples**.

It is important to choose few (3 up to 5, if possible) representative examples for each CM that will adequately document the meaning of the German verb sense. The ParaCrawl Corpus consists of automatically crawled web texts, therefore, filtering the sentences is important. The selected sentences shouldn't be too long or complicated and they should not include inappropriate content. Please look for grammatically correct sentences.

The annotation procedure is simple.

Use the **Add to Lexicon** button when staying on **the selected CM** and on **the selected example sentence**, this will assign this example sentence to the appropriate CM sense and mark it with an asterisk (Fig. 19).

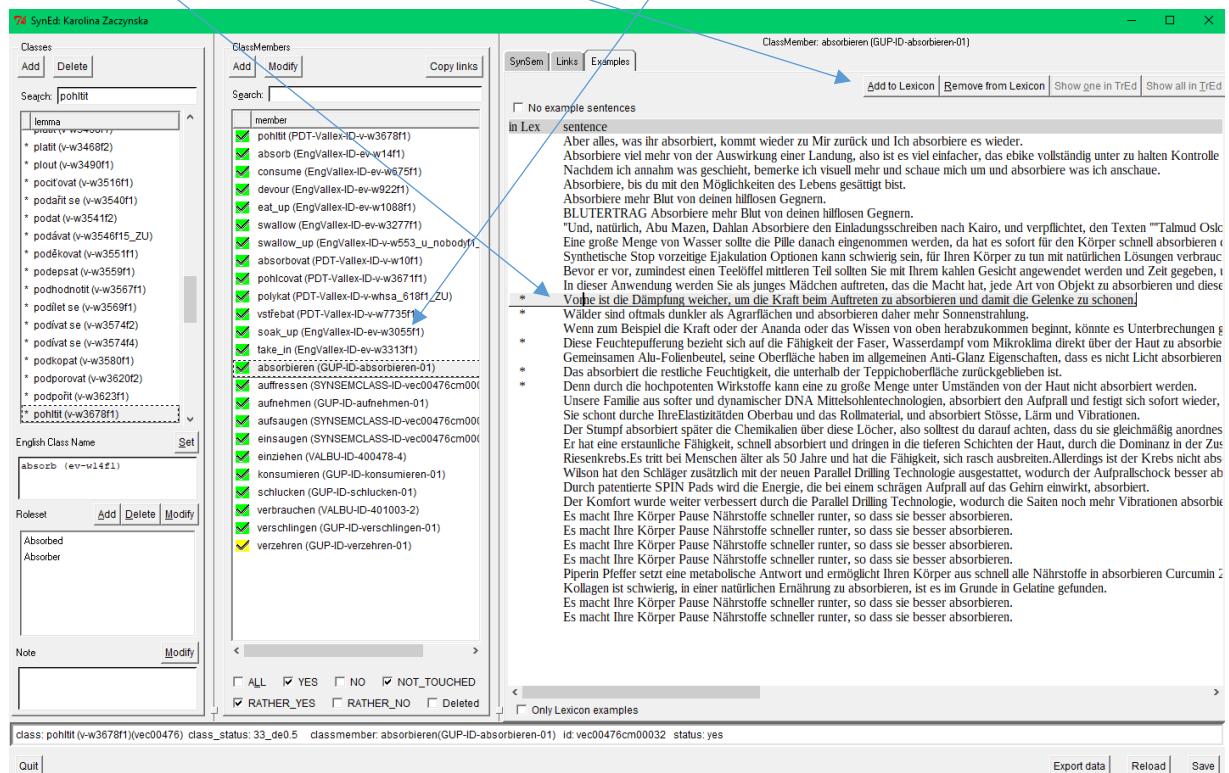


Fig. 19. Assignment of the selected example sentences to *absorbieren* of *pohltit/absorb* class

In case there are no representative examples for the given CM, you check **No example sentences** box.



THE ÚFAL/CKL TECHNICAL REPORT SERIES

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 Hajičová, 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 Katerina 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 Panevová, 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 Panevová, 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 Panevová, 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 Panevová, Jiří Semeka, 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 Ngụ 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 anotátoršké 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 deverbativních 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)*

ÚFAL TR-2019-63 Ján Faryad,
Identifikace derivačních vztahů ve španělštině

ÚFAL TR-2020-64 Marie Mikulová, Jan Hajič, Jiří Hana, Hana Hanová, Jaroslava Hlaváčová, Emil Jeřábek, Barbora Štěpánková, Barbora Vidová Hladká, Daniel Zeman,
Manual for Morphological Annotation. Revision for Prague Dependency Treebank – Consolidated 2020 release

ÚFAL TR-2021-65 Rudolf Rosa, *Technická zpráva o vývoji projektu THEaiTRE v roce 2020*

ÚFAL TR-2021-66 Anna Nedoluzhko, Michal Novák, Martin Popel, Zdeněk Žabokrtský, Daniel Zeman
Coreference meets Universal Dependencies – a pilot experiment on harmonizing coreference datasets for 11 languages

ÚFAL TR-2021-67 THEaiTRobot 1.0, David Košťák, Daniel Hrbek, Rudolf Rosa, Ondřej Dušek,
AI: When a Robot Writes a Play

ÚFAL TR-2021-68 Markéta Lopatková, Václava Kettnerová, Anna Vernerová, Eduard Bejček, Zdeněk Žabokrtský
Valenční slovník českých sloves VALLEX

ÚFAL TR-2021-69 Niyati Bafna, Jan Bodnár, Lukáš Kyjánek, Emil Svoboda, Madga Ševčíková,
Jonáš Vidra, Zdeněk Žabokrtský
Towards Universal Segmentations: Survey of Existing Morphosegmentation Resources

ÚFAL TR-2021-70 Zdeňka Urešová, Eva Fučíková, Jan Hajič, Karolina Zaczynska,
Annotation guidelines for German verbal synonyms included in SynSemClass Lexicon