Articles

Extending Coverage of a Lexicon of Discourse Connectives Using Annotation Projection
Jiří Mírovský, Pavlína Synková, Lucie Poláková

Reflexives in the VALLEX Lexicon: Syntactic Reflexivity and Reciprocity
Václava Kettnerová, Markéta Lopatková, Anna Vernerová

Instructions for Authors

© 2021 PBML. Distributed under CC BY-NC-ND.
Extending Coverage of a Lexicon of Discourse Connectives
Using Annotation Projection

Jiří Mírovský, Pavlína Synková, Lucie Poláková

Charles University, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics

Abstract

We present a method for extending coverage of the Lexicon of Czech Discourse Connectives – CzeDLex – using annotation projection. We take advantage of two language resources: (i) the Penn Discourse Treebank 3.0 as a source of manually annotated discourse relations in English, and (ii) the Prague Czech–English Dependency Treebank 2.0 as a translation of the English texts to Czech and a link between tokens on the two language sides. Although CzeDLex was originally extracted from a large Czech corpus, the presented method resulted in an addition of a number of new connectives and new types of usages (discourse types) for already present entries in the lexicon. We classify and elaborate on reasons why the rest of automatically pre-selected candidates were excluded from the process, and give examples of actual new additions.

1. Introduction

A growing interest in text coherence-aware methods can be traced in many areas of natural language processing (NLP), including tasks such as machine translation (Xiong et al., 2019; Meyer and Webber, 2013), text generation (Kiddon et al., 2016), summarization (Zhang, 2011), information extraction, opinion mining (Turney and Littman, 2003), coherence evaluation (Rysová et al., 2016), or machine translation evaluation (Bojar et al., 2018). Many of these tasks incorporate discourse parsing in text pre-processing and, naturally, discourse parsing methods have come into focus of the discourse research community, including two CoNLL shared tasks (Xue et al., 2015, 2016).

Discourse parsing methods can strongly benefit from two types of language resources – text corpora manually annotated with discourse relations and lexicons of
discourse connectives. Discourse-annotated corpora date back to the Penn Discourse Treebank (PDTB; Miltsakaki et al., 2004) and the RST Discourse Treebank (Carlson et al., 2002), representing two dominant theoretical approaches to discourse coherence representation in text corpora – local (shallow) vs. global discourse models. Both approaches have been later followed for many languages.¹

Electronic lexicons of discourse connectives – as an invaluable resource for both theoretical discourse research and automatic discourse processing – also date back almost two decades: an XML-based and machine readable DiMLex for German (Stede, 2002) and a more human-oriented DPDE, a dictionary of Spanish discourse markers (Briz et al., 2003). Since then, their number has been steadily rising, recently (since 2014) in connection with the COST Action TextLink, dedicated to discourse resources and representations: LexConn for French (Roze et al., 2012), LICO for Italian (Feltracco et al., 2016), CzeDLex for Czech (Mirovský et al., 2017), DiMLex-Eng for English (Das et al., 2018), LDM-PT for Portuguese (Mendes et al., 2018), and others. Most of these resources have been gradually incorporated in Connective-Lex (Stede et al., 2019), a multi-language database of discourse connectives currently covering 10 languages.²

Lexicons of discourse connectives gather and organize structured information about discourse connectives. Discourse connectives are words or phrases explicitly signalling discourse relations, i.e. semantico-pragmatic relations between two text spans (often called arguments). These relations can be either intra- or inter-sentential (i.e. they can occur within one sentence or between two or more sentences). Example 1 from the PDTB exhibits an intra-sentential discourse relation of discourse sense *Comparison.Concession* expressed with connective *though*.³

(1) **Though** Mrs. Thatcher has pulled through other crises, supporters wonder if her steely, autocratic ways are the right formula today. *(PDTB)*

We distinguish two types of discourse connectives, primary and secondary (Rysová and Rysová, 2014). Primary connectives form an almost closed set of mainly one-word expressions belonging mostly to conjunctions (*but, or, however*), particles (*only, too*) and adverbs (*later, previously*).⁴ Secondary connectives belong to an open set of a broad range of expressions that are not yet fully stable or grammaticalized (for these reasons, *the main condition is, that is why*); they can be a part of the sentence syntac-

¹ A summarizing list of discourse-annotated corpora for different languages and within different frameworks can be found at [http://www.textlink.ii.metu.edu.tr/corpus-view](http://www.textlink.ii.metu.edu.tr/corpus-view)

² [http://connective-lex.info/](http://connective-lex.info/)

³ Discourse relations can be expressed by a connective (we call them explicit discourse relations), or understandable only from the context and the meaning of the arguments (we call them implicit).

⁴ According to the traditional Czech word class categorization, particles form an autonomous category. In contrast to adverbs they do not participate in the sentence structure.
tic structure or even stand as a separate clause. Secondary connectives correspond roughly to alternative lexicalizations (AltLexes) in the PDTB terminology.

For using lexicons of discourse connectives in NLP tasks, it is crucial that the lexicons carry linguistic information not only about syntactic properties of the connectives, but most importantly also about their semantic properties, i.e. a list of discourse senses\(^5\) the connective can express and the semantics of the discourse relation arguments (e.g., for the reason-result relation, which of the arguments represents the “reason” and which the “result”). Some of the lexicons were built with this principle from the start (LexConn, CzeDLex), some others were enriched with semantic information in their recent versions (DiMLex, Scheffler and Stede, 2016). All lexicons to be added to Connective-Lex are required to carry the semantic information.

Various strategies may be employed to build electronic lexicons of discourse connectives, depending on available resources – traditional printed lexicons may be consulted, discourse-annotated corpora may be used to extract lexicon data, various projection methods may be used to utilize existing discourse-related resources in another language, etc. However, in any case, building a lexicon with richly annotated entries requires a lot of (subsequent) manual work.

The Lexicon of Czech Discourse Connectives, CzeDLex (Mírovský et al., 2017), was originally extracted from a large Czech discourse-annotated treebank – the Prague Discourse Treebank 2.0 (PDiT; Rysová et al., 2016). The extraction from this 50-thousand-sentences corpus with more than 20 thousand annotated explicit discourse relations produced a lexicon with approx. 200 entries, which have been gradually manually edited since, leading to several published versions of the lexicon. CzeDLex 0.6, published in December 2019 (Synková et al., 2019), contained 204 entries, out of which 76 entries (covering more than 90% of the discourse relations annotated in PDiT) were fully manually checked and supplemented with additional linguistic information. It was the last version of CzeDLex containing solely entries originating in PDiT.

The present article elaborates on theoretical and practical aspects of the subsequent enrichment of the lexicon by exploiting the method of annotation projection and two additional resources – the Penn Discourse Treebank 3.0 (PDTB; Prasad et al., 2019) and the Prague Czech–English Dependency Treebank 2.0 (PCEDT; Hajič et al., 2012a).

Annotation projection is a well established and widely used method of automatic or partially automatic cost-effective linguistic annotation. The purpose of the projection is to induce annotation of a certain language phenomenon in a target language, using an already existing annotation of the phenomenon in a source language and parallel texts/corpora in the two languages.

\(^5\) Throughout this article, the term (discourse) senses is used in compliance with the PDTB terminology when speaking about the senses/meanings of English, PDTB-style-annotated discourse relations, whereas the term discourse (semantic) types refers to the same notion in Czech annotations both in PDiT and in CzeDLex.
The method has been employed in various types of tasks, ranging from morphology to syntax and to semantics. To name just a few examples out of many, Yarowsky and Ngai (2001) used annotation projection for part-of-speech tagging and detection of noun phrases, with English as the source language and French and Chinese as the target languages. Hwa et al. (2005) trained dependency syntax parsers for Spanish and Chinese on data obtained by a projection of manual syntactic annotation in English. Padó and Lapata (2009) exploited possibilities of annotation projection from English to German on the task of semantic roles labeling.

Annotation projection is not unheard of either in the field of discourse annotation: Versley (2010) used annotation projection to induce detection of discourse connectives in German using English–German parallel texts and an automatic discourse parser on the English side. In 2017, Laali and Kosseim studied possibilities of projecting annotation of discourse relations from English to French, creating a discourse annotated French corpus of Europarl data. Sluyter-Gäthje et al. (2020) even used automatically translated texts of the PDTB (and annotation projection) to create a German discourse-annotated corpus, GermanPDTB.

The rest of the article is organized as follows. We describe our method and data in detail in Section 2 and analyze the results in Section 3. Section 4 concludes the article.

2. Data and Methodology

The possibility to use annotation projection to enrich the Lexicon of Czech Discourse Connectives, CzeDLex, with additional data extracted from another discourse-annotated treebank comes from a unique situation of having two key resources at our disposal, the PDTB 3.0 and the PCEDT 2.0:

**PDTB 3.0:** The Penn Discourse Treebank 3.0 (Prasad et al., 2019) is a corpus of English newspaper texts annotated manually with discourse relations. The texts consist of approx. 50 thousand sentences of the Wall Street Journal section of the Penn Treebank (Marcus et al., 1995) and the annotation contains approx. 40 thousand discourse relations of various kinds (including implicit relations and entity-based relations). For our purposes, we have used approx. 26 thousand relations explicitly expressed by a connective.

It is worth noting that PDiT (the original source corpus for CzeDLex)\(^6\) and the PDTB are comparable in genre (journalism), size (50k sentences) and are similar also in the annotation scenario\(^7\) and the extent of the annotated explicit discourse relations (21 thousand vs. 26 thousand).

---

\(^6\) the Prague Discourse Treebank 2.0 (PDiT; Rysová et al., 2016)

\(^7\) although there are differences in the sense hierarchies and e.g. implicit relations were not annotated in PDiT
PCEDT 2.0: The Prague Czech–English Dependency Treebank 2.0 (PCEDT; Hajič et al., 2012a, Hajič et al., 2012b) is a corpus of English–Czech parallel texts and their analysis on several layers of language description in the same annotation scenario as PDiT. Importantly, the English part of the PCEDT contains the same texts as the PDTB, i.e. the Wall Street Journal section of the Penn Treebank (PTB). The Czech part is based on human translations of the English texts to Czech, by design 1:1 sentence-aligned, with an additional automatic alignment on the word/node level on all annotation layers.

Methods used in the research described in the present article were implemented in the Prague Markup Language data format and application framework (PML; Pajas and Štěpánek, 2008), which is a primary format for PDiT, CzeDLex and the PCEDT. From a previous research, also the PDTB (mapped onto the PTB) was available in the PML format (Mírovský et al., 2016). The Prague Markup Language is an XML-based format and application framework designed for multi-layer linguistic annotations with available tools allowing for complex linguistic studies: tree editor TrEd8 for browsing and editing PML data, btred for applying Perl scripts to the data and Prague Markup Language - Tree Query (PML-TQ; Pajas and Štěpánek, 2009) as a powerful, graphically oriented query system.9

The method for CzeDLex enrichment consisted of the following distinctive steps:

1. projection of the PDTB discourse annotation to the Czech part of the PCEDT (PCEDT-cz), see Sec. 2.1 below,
2. transformation of the PDTB discourse senses to the Prague taxonomy (Sec. 2.2),
3. extraction of czedlex-pcedt-cz, a raw PCEDT-cz-based lexicon of connectives (Sec. 2.3),
4. identification of connectives and discourse senses not present in CzeDLex, manual selection of the relevant ones (Sec. 2.4 and 3),
5. merging the selected new data into CzeDLex (Sec. 2.5),
6. manual fixes/annotation (an ongoing work).

We describe the individual steps from the technical point of view in Sections 2.1 – 2.5. Section 3 offers a detailed analysis of the manual selection in step 4 from a linguistic point of view. Step 6 represents an ongoing work, to be finished by the end of the year by a publication of a new version of CzeDLex.

8 https://ufal.mff.cuni.cz/tred/
9 See Mírovský et al. (2016) and Mírovský et al. (2014) for a demonstration how to search with the PML-TQ in the PDTB and PDiT, respectively.
2.1. Annotation Projection

The projection of the discourse annotation from the PDTB to the PCEDT-cz consisted of several sub-steps. The discourse annotation was first mapped from the raw texts to the Penn Treebank phrase structure trees using procedures and the framework described in Mírovský et al. (2016), newly adapted to the annotation scheme of version 3 of the PDTB. Among other things, the adaptation involved a computation of so called GORN addresses, used to map text spans defined by character offsets in the raw texts to nodes in the trees of the Penn Treebank. After the mapping, all attributes of the discourse relations related to text spans became represented by minimal sets of nodes in the PTB, and the relations themselves became represented as links (arrows) between the sets of nodes corresponding to the discourse relation arguments.

Second, the discourse annotation was copied from the Penn Treebank phrase structure trees to the dependency trees of the tectogrammatical layer of the English part of the PCEDT (PCEDT-en), using 1:1 correspondence between terminal nodes of the Penn Treebank and nodes at the analytical (surface syntax) layer of the PCEDT-en (a-nodes), and then links from nodes on the tectogrammatical (deep syntax) layer of the PCEDT-en (t-nodes) to the a-nodes. The annotation obtained at this point was structurally close to the one in PDiT.

Finally, the discourse annotation was projected from English to Czech, i.e. from the PCEDT-en tectogrammatical trees to the PCEDT-cz tectogrammatical trees, using an automatic alignment of nodes on the corresponding t-layers. Errors originating from the automatic alignment form a large part of errors in the projected data and are discussed in Section 2.3 and also in Section 3.

2.2. Sense Taxonomy Transformation

The PDTB and PDiT use similar sets of senses/discourse relation types. Table 1 shows the mapping of the PDTB 3.0 senses to PDiT discourse types used in the transformation. The mapping is not entirely 1:1 – in cases when a single PDTB sense maps to two PDiT discourse types, the more frequent one was used (listed first in the table). Please note that only the sense Expansion.Level-of-detail distinguishes argument semantics in the table, as it maps to two different discourse types in Czech (specifica-
J. Mirovský et al. Extending CzeDLex with Annotation Projection (5-26)

<table>
<thead>
<tr>
<th>PDTB 3 sense</th>
<th>PDiT 2 discourse type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comparison.Concession</td>
<td>concession, opposition</td>
</tr>
<tr>
<td>Comparison.Concession+SpeechAct</td>
<td>pragmatic opposition</td>
</tr>
<tr>
<td>Comparison.Contrast</td>
<td>confrontation</td>
</tr>
<tr>
<td>Comparison.Similarity</td>
<td>conjunction</td>
</tr>
<tr>
<td>Contingency.Cause</td>
<td>reason–result</td>
</tr>
<tr>
<td>Contingency.Cause+Belief</td>
<td>explication</td>
</tr>
<tr>
<td>Contingency.Cause+SpeechAct</td>
<td>pragmatic reason–result</td>
</tr>
<tr>
<td>Contingency.Condition</td>
<td>condition</td>
</tr>
<tr>
<td>Contingency.Condition+SpeechAct</td>
<td>pragmatic condition</td>
</tr>
<tr>
<td>Contingency.Negative-cause</td>
<td>-</td>
</tr>
<tr>
<td>Contingency.Negative-condition</td>
<td>condition</td>
</tr>
<tr>
<td>Contingency.Purpose</td>
<td>purpose</td>
</tr>
<tr>
<td>Expansion.Conjunction</td>
<td>conjunction, gradation</td>
</tr>
<tr>
<td>Expansion.Disjunction</td>
<td>disjunctive alternative, conjunctive alt.</td>
</tr>
<tr>
<td>Expansion.Equivalence</td>
<td>equivalence</td>
</tr>
<tr>
<td>Expansion.Exception</td>
<td>restrictive opposition</td>
</tr>
<tr>
<td>Expansion.Instantiation</td>
<td>instantiation</td>
</tr>
<tr>
<td>Expansion.Level-of-detail.Arg1-as-detail</td>
<td>generalization</td>
</tr>
<tr>
<td>Expansion.Level-of-detail.Arg2-as-detail</td>
<td>specification</td>
</tr>
<tr>
<td>Expansion.Manner</td>
<td>-</td>
</tr>
<tr>
<td>Expansion.Substitution</td>
<td>correction</td>
</tr>
<tr>
<td>Temporal.Asynchronous</td>
<td>precedence–succession</td>
</tr>
<tr>
<td>Temporal.Synchronous</td>
<td>synchrony</td>
</tr>
</tbody>
</table>

| Table 1. The PDTB – PDiT sense transformation table. |

...tion and generalization). Argument semantics of other asymmetric senses projected to the PDiT taxonomy is captured in the direction of the discourse relation (which is represented as a link and depicted by an arrow in tectogrammatical trees).13

2.3. Extraction of a Raw Lexicon

A raw version of CzeDLex was originally extracted from PDiT. The extraction script used a flat list of connectives occurring in the annotated data, manually pre-grouped in the sense of a connective and its variants, modifications and complex forms. The script took the flat list of grouped connectives, went through the annotated discourse data and integrated information from each discourse relation into the raw lexicon, gradually creating entries for individual connectives and their possible discourse types.

---

13 Argument semantics specifies roles of two arguments of an asymmetric discourse relation – e.g., for discourse type reason–result, it specifies which of the arguments is the reason and which one is the result.
Additional information was also collected by the script, such as argument semantics, numbers of occurrences, corpus examples, examples of non-connective usages and others. The extraction process was described in detail in Synková et al. (2017). Manual annotation and corrections then started on these automatically extracted data (the whole process was summarized in Mírovský et al., 2017).

Similar scripts were now used to automatically extract raw lexicon data from the PDTB discourse annotation projected to the PCEDT-cz (we only used relations of types Explicit, AltLex a AltLexC, i.e. relations originally expressed by a connective, with a non-empty counterpart in the word alignment). The flat list of connectives appearing in the PCEDT-cz contained almost 3 thousand entries, most of them representing errors in the word alignment between the English and Czech parts of the PCEDT. We first cut off all single occurrences (over 2 500 entries). From the remaining slightly over 500 entries, the most obvious word alignment errors were deleted, the rest of entries were pre-grouped in the sense of modifications etc. The resulting grouped flat list was used to automatically extract a raw lexicon czedlex-pcedt-cz from the PCEDT-cz, containing over 200 entries (connectives) along with their variants, possible discourse types, complex forms, modifications, examples (original English and Czech translations), corpus counts etc.

2.4. Automatic Pre-Selection and Manual Selection

The extracted czedlex-pcedt-cz was automatically compared with the current version of CzeDLex to mark connectives not appearing in CzeDLex and – for connectives already present in CzeDLex – to mark discourse types not appearing at the respective entries in CzeDLex. This marking produced a list of 92 potential candidates for new connective entries and further 250 potential new discourse types to be added to existing entries.

These candidates were subsequently inspected by two experienced annotators who were asked to mark each candidate (a whole entry or a discourse type) with one of three options meaning USE, POSSIBLY USE and DO NOT USE. The annotators considered the automatically collected examples and (if needed) their broader textual context both in Czech and in original English, for complex cases they entered comments and discussed their choices. This process significantly narrowed the selection of candidates and is analyzed in detail in Section 3.

Only candidates that were marked at least by one of the annotators as USE or by both as POSSIBLY USE (in total, 25 new whole entries and 17 new discourse types for already existing entries) were then actually selected for an inclusion into CzeDLex.

2.5. Merging

The selected 25 connectives and additional 17 discourse types were merged to the current version of CzeDLex, being still subjects to a later detailed manual inspection
and annotation (and even possible eventual deletion) just like any previous CzeDLex entry/discourse type. A new attribute source marking external source was added to the data scheme and filled with value PCEDT for the new data. In the graphical environment, the external source is clearly visible, distinguishing the new data from the original ones. New discourse types of already existing connectives are sorted at the end of possible discourse types of a connective and their counts in the source corpus are not added to the overall count of the connective (otherwise they would disrupt percentages of various discourse types for the connective).

3. Analysis of the Projected Data

This section of the article addresses the process of manual inspection and evaluation of the extracted and pre-selected lexicon data by the annotators. We describe a set of connective and discourse type candidates (out of the automatically pre-selected 92 and 250, resp.) that were in the end not included in CzeDLex and categorize reasons for their exclusion (Section 3.1). Then we present the set of included new connectives and discourse types (Section 3.2). The discussed categories are accompanied by corpus examples, i.e., Czech translations from the Czech part of the PCEDT and the English PDTB originals. Connectives in the examples are highlighted in bold.

3.1. Candidates Not Included in CzeDLex

The reasons for not including some of the pre-selected candidates to CzeDLex can be divided into three main groups; we address them below in detail:

1. differences in the annotation schemes and strategies (Sec. 3.1.1),
2. issues coming from the translation (Sec. 3.1.2),
3. errors originating in the projection process (Sec. 3.1.3).

3.1.1. Differences in the annotation schemes

The issues arising from the differences in the annotation schemes for English (the PDTB approach) and Czech discourse relations (the PDiT approach) included namely the following: differences in distinguishing across discourse senses/types and in definitions of individual senses/discourse types (different annotation guidelines for senses/discourse types with the same label), and differences in the evaluation of individual expressions with respect to actually fulfilling the function of a connective (in a given context). Altogether, these reasons account for approx. 40% of the excluded candidates; the majority of these cases represent differences in the semantic label taxonomies and annotation strategies.

The semantic taxonomies of the PDTB 3.0 and PDiT 2.0 differ first in the presence of senses Expansion.Manner and Comparison.Similarity, which the PDiT approach con-
siders to be rather a part of the syntactic analysis\textsuperscript{14} or a specific case of conjunction (Comparison.Similarity), and second, in the absence of the relation of gradation (a part of Expansion.Conjunction in the PDTB), explication (a part of Contingency.Cause and Contingency.Cause+Belief) and conjunctive alternative (a part of Expansion.Disjunction). Mismatches caused by these differences are complemented by cases where our annotators did not agree with the PDTB interpretation of the given relation.

Example 2 shows a PDTB relation Expansion.Manner.Arg2-as-manner that is not considered to be a discourse relation in PDiT; Example 3 shows a PDTB relation with the same label that would be interpreted as conjunction in the PDiT approach, and Example 4 illustrates a context in which we do not agree with the PDTB interpretation – this context in our opinion contains an Expansion.Equivalence relation, not Expansion.Conjunction.

(2) Potom jsem si všimla, že se auto pohybuje nahoru a dolů, jako kdyby na něm někdo skákal.
[Then I noticed the car was bouncing up and down as if someone were jumping on it.]

(3) Ale firma Honda letos model Accord zrenovovala a udělala z něj vůz střední velikosti.
[But this year, Honda has revamped the Accord and made it a midsized car.]

(4) Podle Cathcartových slov to bude ve společnosti Kidder v nadcházejících letech „hučet jako v úle“. Neboli, jak říká Carpenter opírající se o své zkušenosti z konzultantské firmy: „Teď jsme připraveni jednat."
[In coming years, Mr. Cathcart says, Kidder is “gonna hum.” Or, as Mr. Carpenter, again drawing on his consulting-firm background, puts it: “We’re ready to implement at this point.”]

Differences in classifying certain words (tokens) as connectives are a reason for excluding, e.g., a comma from the set of new connectives, since a comma has in our opinion too many other functions to be used as a reliable signal of a discourse relation,\textsuperscript{15} the adverb nyní [now] was excluded for being considered a part of a bridging relation and a semantic constituent of the sentence rather than a connective. A specific case is represented by non-finite verb structures where the verb form itself is consid-

\textsuperscript{14} The syntactic label Expansion.Manner in the underlying Czech syntactic annotation was not assessed to hold analogically in discourse annotation, as the possibilities of its expression in an inter-sentential setting seemed quite restricted in Czech and most similar cases were judged quite satisfactorily as specification.

\textsuperscript{15} In the PDiT approach, only a colon, a semicolon and a dash are considered to be connectives.
tered both a part of an argument and a connective in the PDTB – cf. the verb *zanechat* [*leaving*] in Example 5.

In the PDIT approach, such (notional) verbs represent the core of a proposition, they are a constitutive part of the argument and thus do not exhibit the main feature of a connective, i.e. being an operator connecting two spans of a text.\(^\text{16}\)

(5) Použití herbicidů by vybilo plodné rostliny a *zanechal* velké pole rostlin se samčí sterilitou, které mohou být opylovány pro získání křížených semen.

[The application of herbicide would kill off the male-fertile plants, *leaving* a large field of male-sterile plants that can be cross-pollinated to produce hybrid seed.]

3.1.2. Issues coming from the translation

Although the human translators of the PCEDT texts from English to Czech were instructed to translate as literally as possible (but fluently), differences originating in the translation are the most common cause for excluding candidates from the CzeDLex enrichment (they account for approx. 50%). During the analysis of the projected data, three main types of differences caused by translation differences were detected.

The most common type was a choice of a more specific Czech connective for a less specific English one, e.g. *as* with a temporal annotation in Example 6 was translated as jelikož [*because*]. In English, *as* is a highly polyfunctional expression (according to the PDTB annotation, it can signal relations from all four major classes of discourse senses). In Czech, and similarly in other languages, a translation of expressions such as *as* necessarily implies a disambiguation among the possible interpretations of the original word.\(^\text{17}\) The Czech translation equivalent in Example 6 jelikož signals the meaning of *reason–result*, it does not have a temporal meaning. Thus, the new temporal meaning of the connective jelikož coming from the PDTB projection had to be excluded from the CzeDLex enrichment.

(6) V Londýně při nestálém obchodování uzavřely akcie níže, jelikož začínající zotavení bylo zeslabeno obchodními výsledky USA, které jsou horší, než se čekalo.

[In London, stocks closed lower in volatile trading as an opening rally was obliterated by worse-than-expected U.S. trade figures.]

---

\(^{16}\) The issue of verb forms representing a connective is a more general one. Surely some verbs have some inner connectivity feature (*imply, cause, mean, contradict, follow...*) but whether to assess them as connectives or as arguments (propositions), or, where to set the border, is a theoretical question for discussion. At present, the Czech annotations and the CzeDLex do not include verbs as connective entries.

\(^{17}\) Disambiguation by translation in general is a well-known topic in translation studies, but also a separate topic in discourse research: disambiguation of (functions of) connectives by their translation, e.g. Meyer (2011). Cartoni et al. (2013).
The second type of “translation reasons” for excluding candidates for CzeDLex enrichment were different properties of English–Czech counterparts both at the word and the higher construction level. A difference at the word level is illustrated in Example 7. English indeed can be in many contexts translated correctly as opravdu,16 but it cannot stand separately at the very beginning of a sentence. Being a constituent within the sentence, it loses its connectivity and becomes a modal particle – thus in the present context, it would be more appropriate to leave it out or to choose a non-literal translation. The connectivity of English indeed is beyond all doubt.

(7) Jeho organická architektura odrážela citlivý vztah k životnímu prostředí již desítky let předtím, než se toto téma stalo populární mezi „rádoby aktivisty“. Wright opravdu celý svůj život tvrdil, že nejvíce se toho naučil studiem přírody.

[Wright’s organic architecture demonstrated a keen sensitivity to the environment decades before it became fashionable among “la-la activists”. Indeed, Wright said all his life that the greatest lessons he learned were derived from the study of nature.]

As for differences at the higher construction level, the most common case was the translation of an English non-finite verb structure by a Czech clause. The English structure by taking... in Example 8 could be translated by a Czech non-finite structure but the sentence would sound unnatural. The chosen translation (když nastoupil...) sounds natural but does not preserve the meaning of the English structure – it is no longer Contingency.Purpose or Expansion.Manner (as annotated in the PDTB), but synchronous or reason–result.

(8) Fromstein upevnil svou kontrolu v dubnu, když nastoupil po Berrym na místo předsedy představenstva.

[Mr. Fromstein solidified his control in April by taking over from Mr. Berry as chairman.]

Apart from the translation by a dependent clause, another option is to translate an English non-finite verb structure by a Czech verbal noun – cf. Example 9 where the structure for loading... was translated as k naložení... where k naložení is a prepositional phrase with the noun in dative.

(9) Sovětské nákupy jsou tak masivní, že vývozci mají potíže sehnat dostatek říčních člunů a vlaků, aby dopravili právě sklzenou středozápadní úrodu do přístavů k naložení na sovětské lodě.

[The Soviet purchases are so massive that exporters are struggling to find

---

16 or, maybe, more precisely as vskutku
enough river barges and trains to move the recently harvested Midwest crop to ports for loading onto Soviet ships.

However, when translated by nouns, text spans forming a discourse argument in the PDTB do not represent an argument in the PDiT approach, as a PDiT argument requires a finite verb as its core\(^{19}\) and these cases were thus excluded from the candidates to CzeDLex enrichment.

Finally, some candidates had to be excluded due to an inadequate translation. In Example 10, the English connective *while* is translated by Czech *místo aby* [instead of], which substantially modifies the sentence meaning. The connective–sense pair (in this context *místo–concession*) is thus unusable.

\begin{example}
Místo aby slíby ohledně velkých zisků rozezněly zvony na poplach, tak toho obvykle nedocílí, částečně proto, že povídačky o tom, jak se dá rychle zbohatnout, se staly pevnou součástí amerického folklóru.
\end{example}

\begin{example}
While the promises of big profits ought to set off warning bells, they often don’t, in part because get-rich-quick tales have become embedded in American folklore.
\end{example}

3.1.3. Projection errors

The least common reason for excluding candidates were projection errors, these cases accounted for approx. 10%. Most of these cases were less obvious errors caused by the automatic word alignment in the PCEDT that had not been detected before in the projection scenario (see Section 2.3). One context with such an error is in Example 11 – although in some other contexts the alignment correctly matched *although* and *i když* as counterparts, in this sentence it picked just the word *i* as the counterpart to *although* (producing a nonsensical connective–sense pair).

\begin{example}
I když se sledovanost v dobách převratných novinek prudce zvýší, v době zklidnění upadá.
\end{example}

\begin{example}
Although viewership soars when big news breaks, it ebbs during periods of calm.
\end{example}

Another example of an error in the word alignment originates in different properties of English and Czech at the structural level. In Example 12 the English connective *yet* is translated by the Czech counterpart *však*, which has a substantially different word order position in this context. The word alignment wrongly picked the more conve-

---

\(^{19}\) This was a practical annotation decision; we are aware of the fact that also non-finite verb structures and deverbative nouns can represent an argument of a discourse relation.
niently positioned adverb *dosud* in the temporal meaning of *so far* as the translation of the connective *yet*, thus (wrongly) associating the relation of *concession* with the temporal connective.

(12) Koloběh bohužel nepřichází ve vlnách, ale v sestupné spirále. Důkazem toho, že dosud nejsme úplně na dně, je však to, že si ještě navzájem nepomáháme.

[Sadly, the cycle appears not as waves but as a downward spiral. *Yet* the evidence that we have not hit bottom is found in the fact that we are not yet helping ourselves.]

### 3.2. Candidates Included in CzeDLex

The list of candidates included to CzeDLex after the manual assessment comprises 25 new whole entries (connectives) and 17 new discourse types for already existing entries. As we wanted to eliminate a possible influence of “translationese” in connective translations, the frequency of new lemmas was checked in a large representative corpus of Czech (Křen et al., 2019) on original Czech texts and for some cases, even professional translators were consulted.

#### 3.2.1. Whole new entries

Primary connectives newly added to CzeDLex as whole new entries are quite rare. They are mostly single-word adverbs and they represent less frequent alternatives to some more common primary connectives. Most of them did not occur in the original PDiT corpus and their existence and connective function was first documented in the PDTB translation, e.g. *kupříkladu* (more commonly *například* [*for example*]), *obdobně* (more commonly *podobně* [*similarly*], see Example 13), *taktéž* (*též, také* [*also, too*]). These new connectives and their more common counterparts are synonyms and have identical discourse functions, although there might be a slight difference in register: the new lemmas appear more formal than the more frequent ones.

(13) Například 88% čtenářů tohoto listu vlastní akcie (což je o něco méně než 91% v obdobném průzkumu loni). [Ale jen 17.5 % uvedlo, že mají na akciovém trhu více než polovinu svých peněz..] **Obdobně** 57% respondentů vlastní podíl v nějakém investičním fondu peněžního trhu a 33% vlastní komunální obligace.

[For example, about 88% of Journal readers owned stock (down slightly from 91% in a similar poll last year). But only 17.5% said they had more than half

---

20 which was actually introduced by the translator, the modification *so far* is not present in the original clause *we have not hit bottom*. 

18
Similarly, 57% of respondents own shares in a money-market mutual fund, and 33% own municipal bonds.\textsuperscript{21}

Secondary connectives are much more frequent as new additions to CzeDLex. In these cases, the core word is mostly a preposition (\textit{namísto} [instead], see Example 14) or a noun (\textit{doba} [time, point], Examples 15 and 16) and the whole connective is a phrase, the exact formulation of which is largely dependent on the chosen way and syntactic possibilities of the translation. In Example 14,\textsuperscript{22} the connective is a phrase with a demonstrative pronoun \textit{toho} (lit. instead of that), but in other contexts it can read: \textit{namísto toho, aby} or \textit{namísto(,) aby} (lit. instead of that/the fact that) – introducing a dependent clause of substitution in Czech. The order of the arguments can switch for different realizations. This fact is reflected in the lexicon by numbering in the discourse type attribute (e.g. correction-1, correction-2, etc.) and by distinguishing the numbered types by further attributes, according to the syntactic structure\textsuperscript{23} that underlies the relation. For newly added secondary connectives, these distinctions are subject to finer manual work.

\begin{itemize}
  \item \text{(14) Pro nové akcie nebyla dosud stanovena žádná cena. \textit{Namísto toho} ponechají společnosti na trhu, ať rozhodne. [No price for the new shares has been set. \textbf{Instead}, the companies will leave it up to the marketplace to decide.]}
\end{itemize}

Connective candidates with the core word \textit{doba} [time, point], or \textit{okamžik} [moment], were added to CzeDLex, as they signal a non-negligible number of temporal relations. Originally, CzeDLex did not include such temporal nouns. They are various phrases containing the core expression, as the pre-processed entry was automatically merged from all instances of this expression within any word chain with a connective function.

For the connective with the core word \textit{doba}, both temporal discourse types have been projected (\textit{synchrony, asynchrony}). For the discourse type of \textit{synchrony}, these phrases include four diverse translations\textsuperscript{24} of the original English connectives \textit{at the same time, at that time, at the time, at that point}, see Example 15. The same amount of translations\textsuperscript{25} is documented for the discourse type of \textit{asynchrony} and the original connectives \textit{ever since, by then, until then, until}, see Example 16. This example nicely

\begin{itemize}
  \item \textsuperscript{21} In this context, the original PDTB sense is \textit{Similarity}, which was transferred to \textit{conjunction} in the Czech taxonomy.
  \item \textsuperscript{22} Like the mentioned primary connectives, \textit{namísto} is also a connective with a more common alternative (\textit{místo} [instead]) and a possible slight shift in register towards formality.
  \item \textsuperscript{23} marked by attributes schema and realizations
  \item \textsuperscript{24} \textit{ve stejné době, v té době, tou dobou, v té samé době}. The last one of them, in our opinion, is in Czech a rather awkward calque of the English \textit{at the same time}.
  \item \textsuperscript{25} \textit{od té doby, co; v té době; do té doby; do doby, než}
\end{itemize}
demonstrates two things: first, the many-to-many translation possibilities of connectives and the effect of the projection in bringing them together, and, the lexicographic challenge in the attempt to systematically capture secondary connectives. For the final record in CzeDLex, this entry will also need a significant manual detailing (schemas, realizations etc.).


[Consider the experience of Satoko Kitada, a 30-year-old designer of vehicle interiors who joined Nissan in 1982. At that time, tasks were assigned strictly on the basis of seniority.]

(16)  Od té doby, co bylo spojení s cholesterolem odhaleno, začali Američané přidávat psyllium do obilovin ke snídani.

[Ever since the link to cholesterol was disclosed, Americans have begun scarfing up psyllium in their breakfast cereals.]

Apart from temporal nouns as core words of new secondary connectives, causative or argumentative nouns as core words extended the list of CzeDLex entries, the original CzeDLex was more reserved in this respect. The newly added noun-based entries include core words srovnání [in/by comparison], předpoklad [assuming, assuming that, providing], kontext [in that context], kontrast [in contrast], následek [as a result, to result in], základ [by, assuming, lit. based on], známka [indication].

Other newly added candidates include adverbs dříve (než) [before, previously, until], původně [originally, previously], skutečně [indeed, in fact], focussing particles předevedím [in particular, especially], zejména [in particular], zvlášt [separately], multiword connectives než aby [rather than] and phrases with prepositions během [while, as], kromě [jiného] [among (other things)]

3.2.2. New discourse types in existing entries

The new 17 discourse types from the projection enriched 12 different already existing CzeDLex entries – five of the connectives were provided with even two new discourse types. An example entry is the connective jak [as, when]. This expression had originally documented four PDiT discourse types (synchrony, asynchrony, reason–result, condition) and a number of non-connective usages as well (e.g. how). The projection from the PDTB revealed two more connective usages, the discourse types of concession, see Example 17, and instantiation, Example 18. These usages may not be very frequent but

26 Some of the newly added connectives are also present in the original PDiT corpus but – for various reasons, incl. simple omission – they were not annotated as connectives before.
they are fully acceptable. According to the sense – discourse type label mapping, the two added relations are identical in both corpora.

(17) Řekl, že sleduje údaje o peněžních zásobách, avšak nepřisuzuje jim prvořadou důležitost, jak navrhují někteří soukromí a vládní ekonomové.
[He said he monitors the money-supply figures, but doesn’t give them paramount importance, as some private and government economists have suggested.]

(18) Taková situace může způsobit spoušť, jak ukázal mimořádný případ, který v Chicagské obchodní komoře nastal toto léto v termínovém obchodu se sójovými boby.
[Such a situation can wreak havoc, as was shown by the emergency that developed in soybean futures trading this summer on the Chicago Board of Trade.]

As for mappings to a different PDiT discourse type, Example 19 with the although connective represents a PDTB Expansion.Exception. The meaning of exception is included in the Czech label restrictive opposition, which covers both exceptions and “milder” restrictive contrasts.

(19) Všichni jsme tu v pořádku, ačkoliv Mame byla nesmírně vystrašená.
[We are all fine here, although Mame was extremely freaked.] 27

4. Conclusion

Coverage (or completeness) of any lexicon is one of its key aspects. We have presented a method for extending coverage of the Lexicon of Czech Discourse Connectives, CzeDLex, using data obtained via annotation projection from a discourse-annotated corpus in English. The process resulted in an inclusion of 25 new full entries and 17 new discourse types for already existing entries.

Translated texts are of a different nature in comparison with texts written originally in a given language. It may be a question of discussion whether it is desirable to expand an original-text-based language resource (a lexicon) by data coming from exploiting translated texts. On the other hand, from the practical point of view, NLP applications using the lexicon should be able to process not only perfect Czech texts but also translated texts and maybe even awkward translations. To address this issue in CzeDLex, we have employed two measures. First, all data originating from English translations are clearly marked as such, and new discourse types for previ-

27 The translation of the English although to Czech in this context was discussed with professional translators. It appears that, while grammatically correct, there are at least two much better translation options to make the sentence sound more natural in Czech.
ously present connectives are kept separately (at the end of the list); it applies also to corpus counts of these discourse types. Second, whenever in doubt, the annotators consulted a large corpus of Czech texts to check expressions that might sound unnatural to Czech native speakers.

All connectives and discourse types added to CzeDLex are subjects to a subsequent detailed check and annotation just like any previous entry. It is therefore possible that before the final publication, some of the new additions will be deleted, merged with another entry or otherwise modified, and in any case supplemented with additional linguistic annotation. The new version of CzeDLex is planned for publication by the end of 2021 in the LINDAT/CLARIAH-CZ repository under the Creative Commons license.

Acknowledgements

The authors gratefully acknowledge support from the Grant Agency of the Czech Republic (project GA19-03490S). The research reported in the present contribution has been using language resources developed, stored and distributed by the LINDAT/CLARIAH-CZ project of the Ministry of Education, Youth and Sports of the Czech Republic (LM2018101). We want to thank Věra Kloudová for consultations on English to Czech translations.

Bibliography


28 https://lindat.cz


Mírovský, Jiří, Lucie Poláková, and Jan Štěpánek. Searching in the Penn Discourse Treebank Using the PML-Tree Query. In Calzolari, Nicoletta, Khalid Choukri, Thierry Declerck,


**Address for correspondence:**
Jiří Mírovský
mirovsky@ufal.mff.cuni.cz
Intitute of Formal and Applied Linguistics
Malostranské nám. 25
118 00 Prague 1
Czech Republic
Abstract

Reflexives, encoding a variety of meanings, pose a great challenge for both theoretical and lexicographic description. As they are associated with changes in morphosyntactic properties of verbs, their description is highly relevant for verb valency. In Czech, reflexives function as the reflexive personal pronoun and as verbal affixes. In this paper, we address those language phenomena that are encoded by the reflexive personal pronoun, i.e., reflexivity and reciprocity. We introduce the lexicographic representation of these two language phenomena in the VALLEX lexicon, a valency lexicon of Czech verbs, accounting for the role of the reflexives with respect to the valency structure of verbs. This representation makes use of the division of the lexicon into a data component and a grammar component. It takes into account that reflexivity and reciprocity are conditioned by the semantic properties of verbs on the one hand and that morphosyntactic changes brought about by these phenomena are systemic on the other. About one third of the lexical units contained in the data component of the lexicon are assigned the information on reflexivity and/or reciprocity in the form of pairs of the affected valency complementations (2,039 on reflexivity and 2,744 on reciprocity). A set of rules is formulated in the grammar component (3 rules for reflexivity and 18 rules for reciprocity). These rules derive the valency frames underlying syntactically reflexive and reciprocal constructions from the valency frames describing non-reflexive and non-reciprocal constructions. Finally, the proposed representation makes it possible to determine which lexical units of verbs create ambiguous constructions that can be interpreted either as reflexive or as reciprocal.
1. Introduction

According to Genuišinė (1987, p. 25), the term reflexive marker, in this paper simply referred to as a reflexive, can be broadly defined as „an element in the verb (affix, ending, etc.) or its environment (particle, pronoun etc.) which has (or once had) a reflexive meaning (of coreference of two semantic roles) as its only or one of many functions“. In various languages, reflexives are involved in a variety of meanings. Due to their high ambiguity, both theoretical and lexicographic description of their functions pose a great challenge. As the meanings encoded by reflexives are primarily associated with changes in the morphosyntactic properties of verbs, they are highly relevant for the description of verb valency.

In Czech, the clitic forms (se, si) and the full forms of reflexives (sebe, sobě, sebou) are available. Both types then occur in various patterns, which are not always easily distinguishable from each other (see esp. Panevová, 2008; Fried, 2007; Kettnerová and Lopatková, 2019). Two main functions of the reflexives can be identified: while the full reflexives have a pronominal function, the clitic reflexives can serve as both a pronoun and a verbal affix cf. German sich (Gast and Hass, 2008) and Polish się (Wiemer, 2007). The pronominal reflexives mark conventionalized syntactic constructions encoding reflexivity (1) and reciprocity (2). The clitic reflexives with the function of the verbal affix are then part of verb forms (3) or of verb lemmas (4).

(1) a. Nenáviděla sebe a své tělo. ‘She hated herself and her body.’
    b. ... a nenáviděl se za to. ‘... and he hated himself for it.’

(2) a. ... a ze všeho nejvíc začnou nenávidět sebe navzájem. ‘... and most of all they start hating each other.’
    b. Byli tu lidé, kteří se dokonce vzájemně nenáviděli, ... ‘There were people here who even hated each other, ...’

(3) a. Platýz se dusil spolu s pórkem zabalený v alobalu v troubě ... ‘The flounder was stewed with leeks wrapped in foil in the oven ... ’

(4) a. Pes se dusil a koule očima. ‘The dog was choking, rolling its eyes.’
    b. Nikdy si nehrála s panenkami, ... ‘She never played with dolls, ...’

1 Further, there is the reflexive possessive pronoun svůj available in Czech. In this paper, we leave it aside as it cannot occupy a valency position of verbs.

2 However, the interpretation of the clitic reflexives as a pronoun is not accepted by some scholars (see Section 3).

3 Unless explicitly stated otherwise, all examples come from the Czech National Corpus, subcorpus SYNv8.
In this paper, we investigate the language phenomena encoded by the reflexive personal pronoun, i.e., reflexivity and reciprocity. Czech – like other Slavic languages – exhibits so-called reflexive-reciprocal ambiguity, i.e., it uses the reflexives for encoding both reflexivity and reciprocity (Nedjalkov, 2007). We introduce the lexicographic representation of these two language phenomena in the VALLEX lexicon, a valency lexicon of Czech verbs, accounting for the role of the reflexives with respect to the valency structure of verbs. This representation makes use of the division of the lexicon into a data component and a grammar component (Lopatková et al., 2016), taking into account that reflexivity and reciprocity are – to a great extent – conditioned by the semantic properties of verbs on the one hand and the systemic nature of the morphosyntactic changes associated with these phenomena on the other.

The paper is structured as follows. Section 2 provides a brief overview of the VALLEX lexicon and the Functional Generative Description, representing its theoretical background. Section 3 sketches approaches to the reflexives se, si in Czech linguistics. Section 4 introduces theoretical findings on reflexivity in Czech and presents its lexicographic representation in the data and the grammar component of the VALLEX lexicon, including basic statistics related to this phenomenon. Section 5 deals in the same way with reciprocity. Finally, reflexive-reciprocal ambiguity in Czech and its lexicographic treatment in VALLEX are described in Section 6.

2. VALLEX Lexicon and Functional Generative Description

Valency, the ability of verbs (and some nouns, adjectives and adverbs) to open a certain number of valency positions for dependent units (here referred to as valency complementations), forms the core of the sentence structure. As the information on valency cannot be inferred on the basis of general rules, it should be systematically described in a lexicon. VALLEX represents such a lexicon, providing a comprehensive description of valency behavior of Czech verbs. VALLEX is theoretically rooted in the valency theory of the Functional Generative Description (FGD; see esp. Sgall et al., 1986; Panevová, 1974–75, 1994).

Valency is a deep-syntactic characteristic of the verb and as such it is captured in FGD on the tectogrammatical layer (the layer of linguistically structured meaning). However, valency has a specific impact on lower layers as well. Two kinds of valency complementations of a verb are distinguished – actants (roughly corresponding to arguments) and free modifications (roughly corresponding to adjuncts). In addition, a third group of so-called quasi-actants is identified, which shares some characteristics with actants and others with free modifications (Lopatková and Panevová, 2006).

Five actants have been identified in FGD: ‘Actor’ (henceforth ACT), ‘Addressee’ (ADDR), ‘Patient’ (PAT), ‘Origin’ (ORIG), and ‘Effect’ (EFF). Actants, corresponding to the surface subject and to direct and indirect objects, are distinguished primarily

---

4 [http://hdl.handle.net/11234/1-3524](http://hdl.handle.net/11234/1-3524)
on a syntactic basis. Free modifications, corresponding to adverbials, are determined on a semantic basis (e.g., DIR3 ‘where to’, TWHEN ‘when’).

All three types of valency complementations can be either obligatory or optional, their obligatoriness tested by the so-called dialogue test (Panevová, 1974–75). Actants and quasi-actants (be they obligatory or optional) and obligatory free modifications form the so-called standard valency frame, characterizing individual lexical units of verbs, i.e., their individual senses. The valency frame consists of a set of valency slots, each corresponding to a single valency complementation. Each slot is characterized by a label marking the relation of the complementation to its governing verb (e.g., ACT or DIR3), by information on its obligatoriness and on possible morphemic forms determining its surface expression.\(^5\)

To capture valency changing phenomena, such as diatheses, reflexivity, reciprocity, and semantic alternations, VALLEX makes use of the division of the lexicon into a data component (providing information specific to individual lexical units of verbs in their active, non-reflexive and non-reciprocal uses) and into a grammar component (capturing regular valency changes related to diatheses, reflexivity and reciprocity in the form of formal rules); see Lopatková et al. (2016). VALLEX thus reflects the fact that valency-related phenomena are typically conditioned by the semantic properties of verbs (and thus must be listed in the form of lexicon entries characterizing their individual lexical units) on the one hand and the fact that morphosyntactic changes brought about by these phenomena are systemic on the other.

In the description of valency changing phenomena, an abstract model of a situation denoted by a lexical unit of a verb, as proposed by Apresjan (1992) and by Mel’čuk (2004), has proved useful. This model presupposes that such a situation is delineated by a certain number and type of semantic (situational) participants that characterize the lexical unit of the verb in a unique way. As no precise metalanguage for their description has been set up, these participants are labeled with intuitively delimited semantic roles. As shown in more detail in Sections 4 and 5, this model makes it possible to identify the mapping between semantics and deep and surface syntax, and thus to pinpoint the main characteristics of changes associated with individual valency changing phenomena (see Kettnerová, 2014, as well).

\(^5\) Morphemic forms of individual actants and quasi-actants are determined by their governing verbs and as such they are not easily predictable even within individual semantic groups of verbs. As a result, they should be listed for each actant and quasi-actant separately. With free modifications, morphemic forms are not listed as they are implied by their semantics.
3. Reflexives in Czech

Reflexivity (in a broad sense) covers all uses of verbs marked by the reflexive. In Czech, the clitic forms *se/si* and the full forms *sebe/sobě/sebou* of the reflexive are available.\(^6\) While the pronominal character of the latter is unquestionable, the status of the clitic reflexives raises controversy. Some scholars deny the pronominal status of the clitic reflexives on the basis of their morphosyntactic behavior (see esp. Karlík, 1999; Oliva, 2000, 2001; Esvan, 1997; Veselý, 2018). For example, the clitic reflexives fuse with the verb forms of the 2nd singular of the past tense (5-a) and of the conditional (5-b), they cannot occur in coordination (5-c) and cannot be used in ellipses (5-d).\(^7\) From this point of view, the clitic reflexives can only be interpreted as verbal affixes.

\[(5)\]

a. *Myl ses už?* (made-up)
   ‘Have you washed yourself already?’

b. *Chces říci, že by ses nebránil, kdyby ti někdo bral, co ti patří?*
   ‘Do you want to say that you would not defend yourself if somebody took what is yours?’

c. *Lehce zranil sebe/*se i spolujezdkyni.*
   ‘He slightly injured himself as well as his co-driver.’

d. *Jaký největší dar jsi mu dala? Sebe./*Se.*
   ‘What is the greatest gift you have given him? Myself.’

Other linguists, however, argue that the clitic reflexives act as verbal affixes in some constructions and as the reflexive personal pronoun in others. The clitic reflexives of the latter type – similar to the full reflexives – mark the referential identity of the valency complementation occupied by the reflexive and another valency complementation, typically the one expressed in the subject position. Those clitic reflexives are then classified as the reflexive pronoun that are substitutable by their respective full forms (see esp. Komárek et al., 1986; Komárek, 2001; Panevová, 2001, 2008). This view is justified by the fact that the choice between the clitic and full form of the reflexive is conditioned by changes in the topic-focus articulation, rather than by a referential difference between them (Fried, 2004, 2007). With respect to changes in the topic-focus articulation, the clitic reflexives can be seen as such feature as well (e.g., compare *Snaží se umýt.* ‘He/she tries to wash himself/herself.’ with *se* belonging to both verbs and the sentence with the full reflexive *Snaží se umýt sám sebe.* ‘He/she tries to wash himself/herself.’). For haplology of the reflexives in Czech see esp. Rosen (2014).

---

\(^6\) Besides verbs, the full forms of the reflexive can also appear with nouns (e.g., *jeho strach o sebe* ‘his worry about himself’), with adjectives (e.g., *hrdý na sebe* ‘proud of himself’) and with adverbs (e.g., *kolmo na sebe* ‘perpendicularly to each other’). The clitic reflexives can sporadically mark also verbal nouns denoting actions (e.g., *stýskání (si)* ‘complaining’), see esp. Veselovská (2001), and adjectives formed from transgressives (e.g., *stýskající (si)* žena ‘the complaining woman’).

\(^7\) Thanks to one of the reviewers for pointing out that haplology, to which the clitic reflexives are typically subject (in contrast to the full forms), can be seen as such feature as well (e.g., compare *Snaží se umýt.* ‘He/she tries to wash himself/herself.’ with *se* belonging to both verbs and the sentence with the full reflexive *Snaží se umýt sám sebe.* ‘He/she tries to wash himself/herself.’). For haplology of the reflexives in Czech see esp. Rosen (2014).
focus articulation, the reflexive pronoun then behaves like other personal pronouns, compare examples (6-a) and (6-b).

(6)  a. Sebe/Jeho/Tebe nenávidím. (made-up)
    b. Nenávidím se/ho/tě. (made-up)
    ‘I hate myself/him/you.’

Here we adopt the latter view as it allows us to treat semantically equivalent constructions that differ only in the form of the reflexive in the same manner.

4. Syntactic Reflexivity

Languages typically have at their disposal some linguistic means for expressing the situation where an entity (typically of an animate and volitive nature) acts on himself/herself (see esp. Faltz, 1985; Frązyngier and Walker, 2000b).\(^8\) This situation can be referred to as a reflexive situation. The reflexive situation thus requires at least a binary predicate where two of its valency complementations, corresponding to two distinct semantic participants, refer to the same entity (which is either singular or plural). The reflexive situation can be exemplified by the scheme in Figure 1 with the verb *vidět* impf ‘to see’.

Conventionalized constructions expressing a reflexive situation are called here syntactically reflexive constructions.\(^9\) In these constructions, the reflexive (of both clitic and full forms) has a pronominal status (see Section 3), marking referential identity of a participant expressed in the valency position occupied by this reflexive personal pronoun and a participant in another valency position.\(^10\) The latter position predominantly involves the subject (henceforth referred to as subject-oriented syntactic reflexivity), see example (7-a). In rare cases, the direct or indirect object position is affected (henceforth object-oriented syntactic reflexivity), see examples (7-b) and (7-c). As a result, the two valency complementations onto which these participants are mapped corefer with each other.

---

\(^8\) König and Gast (2008) delimit reflexivity on the basis of the so-called reflexive predicate, i.e., a binary predicate R acting on a set A and meeting the following formula:

\[ \forall x \in A \ [ R(x, x) ] \]

Further, they extend the notion of reflexivity to predicates with more than two arguments, two of which are instantiated by the same participant.

\(^9\) See also footnote 23 for the distinction between the reflexivization as a syntactic operation and the reflexivization as a word formation process.

\(^10\) This issue has been extensively discussed esp. within the binding theory proposed by Chomsky (1981) and further elaborated, e.g., by Reinhart and Reuland (1993) and by Reuland and Everaert (2001). Within the principles and parameters framework, reflexivity – with respect to the role of the lexicon in the linguistic description – has been discussed by Reinhart and SILONI (2005) as well.
Figure 1. The mapping of semantic participants, valency complementations and surface positions in an example of syntactic reflexivity affecting ACT and PAT of the verb *vidět* impf: Petr\_ACT se\_PAT viděl (v zrcadle).

≈ Petr\_ACT viděl sám se\_PAT (= Petra) v zrcadle.

‘Peter saw himself in the mirror.’

(7) a. To, že muž na sebe doma nedbá, je jedním ze zdrojů problémů v partnerském soužití.

‘The fact that a man does not care about himself at home is one of the sources of problems in partners’ lives.’

b. Ale ochraňoval ji i před sebou samou.

‘But he also protected her from herself.’

c. Jana cítí, jak mu na sobě samém záleží. (modified)

‘Joan feels how much he cares about himself.’

Syntactically reflexive constructions result from the syntactic operation of *syntactic reflexivization*, making use of the reflexive personal pronoun as its primary linguistic marker. The operation is applied to the valency frames underlying non-reflexive constructions and it affects pairs of valency complementations (corresponding to two separate but referentially identical semantic participants). The changes in the lexical and morphosyntactic expression of the valency complementations involved in reflexivity are regular enough to be described by formal rules (see Section 4.2).

The changes in the lexical expression of valency complementations concern the use of the reflexive personal pronoun. The morphosyntactic changes are restricted to the agreement of predicative complements (referred to as complements or verbal attributes in FGD) in syntactically reflexive constructions marked by the clitic reflexive *se*. In these constructions, the predicative complement typically agrees with the nominative subject, as exemplified in (8-a) with the predicative complement *jako outsider* ‘as an outsider’ agreeing with the pro-dropped subject. In limited cases, the predicative complement can agree with the accusative clitic reflexive *se* as well (i.e., the
agreement is the same as for the full reflexive *sebe*). Compare (8-b) with the predicative complement *jako studenta* ‘as a student, agreeing with the direct object expressed by the accusative clitic reflexive *se*, and its modification (8-c) with the full reflexive in the accusative *sebe*, exhibiting the same agreement.\(^{11}\)

\[(8)\]

\begin{enumerate}
\item a. *Když byla menší, viděla se\textsubscript{acc} jako outsider\textsubscript{nom} …*  
‘When she was smaller, she saw herself as an outsider …’
\item b. *V nejranější revoluční vzpomínce z následujících dní se\textsubscript{acc} vidím jako studenta\textsubscript{acc} 1. ročníku gymnázia …*  
‘In the earliest revolutionary memory of the following days, I see myself as a first-year grammar school student …’
\item c. *V nejranější revoluční vzpomínce z následujících dní vidím sebe\textsubscript{acc} jako studenta\textsubscript{acc} 1. ročníku gymnázia …* (modified)
\end{enumerate}

4.1. Syntactic Reflexivity in the Data Component

In the data component of the *VALLEX* lexicon, the attribute reflex, which is assigned to individual lexical units of verbs, captures the information on syntactic reflexivity in the form of pairs of the valency complementations affected by reflexivity (this applies to 2,039 lexical units in total, represented by 2,291 verb lemmas in *VALLEX*). This information is limited to actants, quasi-actants and obligatory free modifications.\(^{12}\)

Subject and object-oriented reflexivity are not explicitly distinguished since this distinction follows from morphemic forms of the valency complementations involved in reflexivity. Subject-oriented reflexivity is much more frequent in the *VALLEX* data than object-oriented reflexivity: there are 2,255 attested pairs of valency complementations allowing for the former type of reflexivity and 16 attested pairs allowing for the latter one.

\(^{11}\) The changes in agreement of predicative complements are often taken as evidence that the clitic reflexive has the role of the verbal affix intransitivizing the verb (e.g., Veselý, 2018).

\(^{12}\) Optional free modifications, standing outside the so-called standard valency frame, can be affected by syntactic reflexivity as well; see, e.g., the sentence *Ostatní vojáci ACT si BEN chystali snídani, balili si BEN věci …* ‘The other soldiers were preparing their breakfast, packing their things …’, in which the actant *ACT* and the optional free modification *BEN* are involved in reflexivity. Although these cases were not annotated (as optional free modifications are not systematically captured in *VALLEX*), we expect that reflexivity affecting optional free modifications is governed by the same principles as reflexivity of actants, quasi-actants and obligatory free modifications.
Examples (9-a) to (9-c) provide the three most frequent pairs of valency complementations involved in subject-oriented syntactic reflexivity, as attested in VALLEX.\textsuperscript{13} In these constructions, one of the valency complementations, typically \textsc{act}, is expressed by the (nominative) subject. The other complementation, being filled by the reflexive personal pronoun, can be expressed on the surface either as the (accusative) direct object (9-a), as an indirect object (9-b), or as an adverbial (9-c). The following prepositionless and prepositional cases are attested with actants and quasi-actants in the data:\textsuperscript{14} 2, 3, 4, 7, do+2, k+3, mezi+4, na+4, na+6, nad+7, o+4, o+6, od+2, po+6, podle+2, pro+4, proti+3, před+4, před+7, při+6, s+7, u+2, v+4, v+6, z+2, za+4, za+7.\textsuperscript{15}

(9) a. \textit{naštovat} \textsuperscript{pf} ‘to make angry’ ... \textsc{act}_1 \textsc{pat}_4
   reflex: \textsc{act}–\textsc{pat}
   
   \textit{Naposledy jsem naštival já} \textsc{act}_sám \textsc{sebe} \textsc{pat}.
   ‘Last time I made myself angry.’

b. \textit{dokazovat} \textsuperscript{impf}–\textit{dokázat} \textsuperscript{pf} ‘to prove’ ... \textsc{act}_1 \textsc{addr}_3 \textsc{pat}_4,\text{\textit{zda,ze,cont}}
   reflex: \textsc{act}–\textsc{addr}
   
   \textit{Češi mají menší potřebu dokazovat si} \textsc{addr}_svou \textsc{svébytnost}.
   ‘Czechs have less need to prove their independence to themselves.’

c. \textit{miřit} \textsuperscript{impf} ‘to aim’ ... \textsc{act}_1 \textsc{pat}_4,\text{\textit{7}} \textsc{dir}_3
   reflex: \textsc{act}–\textsc{dir}_3
   
   \textit{Pistolí miřil střídavě (na sebe, na policistu a zase na sebe)} \textsc{dir}_3, ...
   ‘He aimed his pistol alternately at himself, at the policeman, and at himself again, ...’

\textsuperscript{13} Eight different pairs of valency complementations that can be affected by subject-oriented reflexivity are contained in the data – four pairs of actants (the left column), one pair of an actant and a quasi-actant (the middle column), and three pairs of an actant and a free modification (the right column):

<table>
<thead>
<tr>
<th>\textsc{act–pat}</th>
<th>\textsc{act–obst}</th>
<th>\textsc{act–dir}_3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,543</td>
<td>3</td>
<td>106</td>
</tr>
<tr>
<td>\textsc{act–addr}</td>
<td>468</td>
<td>\textsc{act–dir}_1</td>
</tr>
<tr>
<td>\textsc{act–eff}</td>
<td>43</td>
<td>\textsc{act–loc}</td>
</tr>
<tr>
<td>\textsc{act–orig}</td>
<td>39</td>
<td></td>
</tr>
</tbody>
</table>

\textsuperscript{14} Morphemic forms of free modifications are not explicitly indicated in VALLEX as they are implied by the semantic type of the complementation, see also footnote 5.

In the VALLEX notation, numbers stand for morphological cases (1=nominative, 2=genitive, 3=dative, 4=accusative, 6=locative, 7=instrumental); in the case of prepositional groups, the preposition precedes the number indicating the respective case (prepositions are not translated here as they can have various interpretations depending on their governing verbs). Further, clausal complementations are marked by the respective subordinating conjunction (e.g., \textit{aby} ‘in order to’, \textit{zda} ‘whether’, \textit{že} ‘that’); the abbreviation \textit{cont} stands for a clausal complementation introduced by an interrogative pronoun or an interrogative adverb.

\textsuperscript{15} As for the frequency of individual morphemic forms, prepositionless cases are attested with 1,519 LUs (1,054 accusatives, 402 datives, 40 instrumentals, and 38 genitives; more than one of these forms can express a single affected valency complementation) and 661 LUs fall into the listed prepositional cases (the most frequent being o+6 (159 LUs), na+4 (158 LUs), and s+7 (110 LUs)).
Object-oriented syntactic reflexivity is rather rare, see examples (10-a) to (10-c), illustrating pairs of valency complementations involved in this type of reflexivity. As for their surface realization, two options occur in the VALLEX data. First, in most cases, one of the valency complementations affected by syntactic reflexivity is expressed as the (accusative) direct object, and the other, which is filled by the reflexive personal pronoun, is realized as an indirect object or as an adverbial, both expressed by a prepositional case. Second, both complementations involved in reflexivity are expressed on the surface as indirect objects: the first one in the dative, and the other, filled by the reflexive personal pronoun, in a prepositional case, forming thus a subjectless construction. The following prepositional cases of the valency complementation occupied by the reflexive pronoun are attested in the VALLEX data: do+2, na+6, o+4, o+6, od+2, proti+3, před+7, s+7, v+4.

(10) a. uchránit pf ‘to protect’ … ACT 1 PAT 4 EFF od+2, proti+3, před+7, aby
reflex: PAT–EFF

Že nás PAT pomocí blažené relaxace uchrání před sebou EFF samými.
‘That they will protect us from ourselves by blissful relaxation.’

b. dovádět impf – dověst pf ‘to bring’ … ACT 1 PAT 4 DIR3
reflex: PAT–DIR3

Meditace mě PAT ale dovedla k sobě DIR3 samému, …
‘Meditation, however, brought me to myself, …’

c. záležet impf ‘to care’ … ACT 3 PAT na+6, aby
reflex: ACT–PAT

Dávejte okolí najevo, že vám ACT na sobě PAT záleží.
‘Show people around you that you care about yourself.’

If a lexical unit of a verb allows more than one pair of its valency complementations to be involved in syntactic reflexivity, more than one attribute reflex is assigned to the lexical unit, distinguished by Arabic numerals. From 2,039 lexical units with the

16 The following pairs are attested in the VALLEX data (four pairs of actants in the left two columns and one pair of an actant and a free modification in the right column):

| PAT–EFF | 6 | ACT–PAT | 3 | PAT–DIR3 | 2 |
| ADDR–PAT | 4 | ACT–ADDR | 1 |

17 VALLEX attests the only verb připomínat impf – připomenout pf for which it is the other way around, i.e., the coreferred valency complementation is expressed as the dative indirect object and the one occupied by the reflexive pronoun as the direct object, see the following example:

reflex: ADDR–PAT

Jejich majitelé chtějí mit doma tvora, který jim ADDR připomíná sebe PAT sama nebo jiného člověka.
‘Their owners want to have a creature at home that reminds them of themselves or of another person.’

18 With 16 LUs allowing for object-oriented reflexivity, the prepositional cases od+2 and před+7 appear with 4 LUs, o+6 and proti+3 with 3 LUs, others are attested for just 1 or 2 LUs.
attribute reflex, 232 can have more than one pair of their valency complementations affected by reflexivity. For example, with the verb *ochraňovat* impf – *ochránit* pf, three pairs of its valency complementations can be affected by syntactic reflexivity. Two of them exemplify subject-oriented reflexivity, see examples (11-a) and (11-b), and one falls under object-oriented reflexivity, see example (11-c).

(11) *ochraňovat* impf – *ochránit* pf ‘to protect’ … ACT 1 PAT 4 EFF od-2, prof-3, pred-7
   a. reflex1 : ACT–PAT
      Před rakovinou se PAT člověk ACT může ochránit z velké části sám.
      ‘A person can himself/herself protect himself/herself from cancer for the most part.’
   b. reflex2 : ACT–EFF
      …, protože on ACT ji před sebou EFF samým ochránit nemůže.
      ‘…, because he cannot protect her from himself.’
   c. reflex3 : PAT–EFF
      … snaha ochránit nás PAT před sebou EFF samými.
      ‘… attempt to protect us from ourselves.’

4.2. Syntactic Reflexivity in the Grammar Component

Two basic rules are formulated for syntactic reflexivity: one for subject-oriented syntactic reflexivity and the other for object-oriented reflexivity. Both these rules, applied to the valency frames of verbs underlying their non-reflexive constructions, allow the valency frames describing their reflexive constructions to be derived. Further, the basic rule for subject-oriented reflexivity is complemented by a supplementary rule, describing changes in agreement of predicative complements.

This supplementary rule is applied to the valency frames resulting from the basic rule.

Each rule consists of three sections (see Figures 2, 3 and 4):

A header identifies the rule.

Conditions indicate properties that a lexical unit of a verb and its valency frame must have so that the rule can be applied. Two properties are relevant here:

- First, the lexical unit must be characterized by the attribute reflex, identifying the pair of the valency complementations affected by reflexivity; these complementations are represented by variables X and Y in the case of subject-oriented reflexivity (with X reserved for the nominative complementation) and by Y and Z in the case of object-oriented reflexivity.
- Second, in the basic rules, the valency complementations X in the case of subject-oriented reflexivity and Y in the case of object-oriented reflexivity are restricted by morphemic forms introduced in their subscripts. Further, the supplementary rule restricts the surface form of the complementation Y
to the reflexives se, sebe and the morphemic forms of Z to the forms that follow the subordinate conjunction jako introducing predicative complements.

A set of actions indicates changes in the valency frames of verbs necessary to generate frames underlying their syntactically reflexive constructions. Two changes are relevant here:

- First, the valency complementation Y in the case of subject-oriented reflexivity and Z in the case of object-oriented reflexivity must be filled by the reflexive personal pronoun in the respective morphemic form prescribed for this complementation: the abbreviation REFL is used for the form of the reflexives specified in the subscript by the variable form (standing for both prepositionless and prepositional cases). Other possible morphemic forms of the respective valency complementation, if present in the valency frame, are deleted. Further, agreement of the predicative complement is changed.
- Second, the rules stipulate that in syntactically reflexive constructions, both affected valency complementations must be obligatorily present in the deep structure. As for the surface structure, X in the subject position can be elided; other affected complementations (Y in subject-oriented reflexivity and both Y and Z in object-oriented reflexivity) must be present in the surface structure of reflexive constructions.

If more than one prepositionless or prepositional cases are listed for the valency complementation Y in the case of subject-oriented reflexivity and Z in the case of object-oriented reflexivity, the respective basic rule is applied repeatedly, generating more than one valency frame.

Rules for subject-oriented reflexivity

<table>
<thead>
<tr>
<th>Subject-oriented syntactic reflexivity</th>
<th>reflex_basic_subj</th>
</tr>
</thead>
<tbody>
<tr>
<td>reflex:</td>
<td>X–Y</td>
</tr>
<tr>
<td></td>
<td>X₁ &amp; Y</td>
</tr>
<tr>
<td>Y</td>
<td>form → REFL_form</td>
</tr>
<tr>
<td></td>
<td>other forms → ∅</td>
</tr>
<tr>
<td>obligatoryness</td>
<td>X, Y</td>
</tr>
</tbody>
</table>

*Figure 2. The basic rule reflex_basic_subj indicating changes in valency frames of verbs in the case of subject-oriented reflexivity.*
The basic rule reflex_basic_subj, given in Fig. 2, captures all cases of subject-oriented reflexivity recorded in the VALLEX data, namely it is applied to 2,255 pairs of valency complementations in 2,032 lexical units in VALLEX.

Let us use one of the lexical units of the verb vnímat impf ‘to perceive’ to exemplify the application of the rule reflex_basic_subj. It is determined in the data component of the lexicon that two of its actants, ACT and PAT, see its valency frame in (12), can be affected by reflexivity, i.e., they can refer to the same entity. As ACT has the nominative form, the reflex_basic_subj rule is employed to generate the valency frame underlying the reflexive construction, with X instantiated as ACT, and Y as PAT. In line with the rule, the accusative PAT is filled by the clitic or full form of the reflexive pronoun in the accusative (se, sebe). The remaining form (the subordinating conjunction že) is then deleted as it cannot appear in syntactically reflexive constructions, see the resulting valency frame in (13) underlying the reflexive constructions. Both ACT and PAT must be present in the deep syntactic structure of reflexive constructions. As for the surface, while ACT can be elided (as Czech is a pro-drop language), PAT must be present on the surface, otherwise the resulting constructions would not have the reflexive meaning, see examples (13-a) and (13-b). The form of EFF, realized on the surface as a predicative complement, is not affected; hence it still agrees with PAT in the accusative. Compare the expression jako dobrodruha a psance ‘as an adventurer and outlaw’ in the example in (12) on the one hand and in (13-a) on the other.

(12) vnímat impf ‘to perceive’ ... ACT 4,że EFF jako+4, jako+adj-4
    reflex: ACT–PAT
    Vnímal ho PAT.acc (jako dobrodruha a psance) EFF.acc. (modified)
    ‘He perceived him as an adventurer and outlaw.’

(13) ACT 1 PAT se,sebe EFF jako+4, jako+adj-4
    a. ... a sám sebe PAT.acc vnímal (jako dobrodruha a psance) EFF.acc.
       ‘... and he perceived himself as an adventurer and outlaw.’
    b. ... chvíle, kdy se PAT.acc vnímám především (jako triatlonistu (byť bývalého)) EFF.acc.
       (Araneum Bohemicum)
       ‘... times when I perceive myself primarily as a triathlete (even a former one).’

Generating the valency frame underlying the reflexive construction with a predicative complement in the nominative requires the application of the supplementary rule reflex_compl_se, given in Fig. 3. This rule is applied to the valency frame (13), which is an output of the basic rule. The PAT with the form se, sebe is identified as Y and EFF with the form jako+4, jako+adj-4 as Z. The rule then stipulates that PAT is restricted to the clitic reflexive se and the forms jako+1, jako+adj-1 are added to the list of possible expressions of EFF, see the resulting valency frame in (14). Compare the nominative
Subject-oriented syntactic reflexivity

<table>
<thead>
<tr>
<th>Reflex:</th>
<th>[X–Y \atop Y \text{se, sebe} &amp; Z \text{jako+4, jako+adj-4}] [\text{EFF[COMPL]}]^{†}</th>
</tr>
</thead>
</table>

\(Y\)  
- se, sebe → se  

\(Z\)  
- jako+4 → jako+1, jako+4  
- jako+adj-4 → jako+adj-1, jako+adj-4

† The rule limits the variable \(Z\) to the complementations \text{EFF} and \text{COMPL} corresponding to predicative complements.

Figure 3. The supplementary rule \text{reflex\_compl\_se} for subject-oriented reflexivity.

Form \textit{jako triatlonista} (byť bývalý) in (14) and the accusative form \textit{jako triatlonistu} (byť bývalého) in (13-b).

(14) \quad \text{ACT}_1 \ \text{PAT} \ \text{se} \ \text{EFF} \ \text{jako+1,jako+4,jako+adj-1,jako+adj-4}  
\quad \ldots \ \text{čvíle, kdy se PAT_{acc} vnímám především (jako triatlonista (byť bývalý))) EFF_{nom.}  
\quad (\text{modified})  
\quad ‘\ldots \text{times when I perceive myself primarily as a triathlete (even a former one).’}

The supplementary rule \text{reflex\_compl\_se} is applied to 39 lexical units in \text{VALLEX}.

Rule for object-oriented reflexivity

Reflexive constructions falling under object-oriented reflexivity are described by valency frames generated by the rule \text{reflex\_basic\_obj}, given in Fig. 4. This rule covers all cases of object-oriented reflexivity recorded in the \text{VALLEX} data, namely it is applied to 16 pairs of valency complementations in 16 lexical units in \text{VALLEX}.

For example, to generate the valency frame underlying the reflexive construction of the verb \text{smiřovat \textit{impf} – smířit \textit{pf}} ‘to reconcile’ in (16), the rule \text{reflex\_basic\_obj} is applied to the valency frame (15) for non-reflexive constructions of the verb, see the example in (15). The rule identifies the accusative \text{PAT} as the variable \(Y\) and \text{ADDR} as \(Z\); the output valency frame is provided in (16).

(15) \quad \text{smiřovat \textit{impf} – smířit \textit{pf}} ‘to reconcile’ … \text{ACT}_1 \ \text{ADDR}_{s+7} \ \text{PAT}_{4}  
\quad \text{reflex:} \ \text{ADDR–PAT}  
\quad \text{Náboženství je cokoliv, co tě PAT smiřuje se světem ADDR. (modified)}  
\quad ‘Religion is anything that reconciles you with the world.’

(16) \quad \text{ACT}_1 \ \text{ADDR}_{se sebou} \ \text{PAT}_{4}  
\quad \text{Náboženství je cokoliv, co tě PAT smiřuje se sebou ADDR samým.}  
\quad ‘Religion is anything that reconciles you with yourself.’
Object-oriented syntactic reflexivity

| reflex:       | Y–Z                      |
|              | $Y_{3,4}$ & Z            |
| Z            | prep+case $\rightarrow$ REFL prep+case |
| obligatoryness | $Y, Z$                   |

Figure 4. The basic rule reflex_basic_obj indicating changes in valency frames of verbs in the case of object-oriented reflexivity.

5. Syntactic Reciprocity

Reciprocal situations are those situations where one entity acts on another entity and, vice versa, the latter acts on the former as well. As with reflexive situations, reciprocal situations require at least binary predicates such that two of their valency complementations correspond to two distinct semantic participants. Unlike reflexivity (where the affected complementations refer to the same entity), the complementations affected by reciprocity refer to two distinct entities (either of which is singular or plural). Further, reciprocity is characterized by complex mapping between semantic participants and valency complementations, as exemplified by the scheme in Figure 5 with the verb nařknout pf ‘to accuse’.

In contrast to reflexivity, reciprocity can be usually expressed in individual languages by diverse means (see esp. Kemmer, 1993; Frazier and Walker, 2000a; Nedjalkov, 2007; König and Gast, 2008; Evans et al., 2011), which gives evidence that reciprocity is less grammaticalized than reflexivity.

In Czech, reciprocity can be syntactically or lexically encoded. In the former case, syntax provides specific constructions conventionalized for expressing reciprocity (henceforth syntactically reciprocal constructions). In the latter case, a lexical unit of a verb contains the feature of reciprocity in its lexical meaning (henceforth inherently re-

---

19 In rare cases, three participants are involved in such a relation. In VALLEX, only two such lexical units of verbs are attested: představovat impf – představit pf and seznamovat impf – seznámit pf ‘to introduce’. Due to their sparsity, we leave these situations aside here.

20 König and Gast (2008, p. 7) describe a reciprocal predicate as a binary predicate $R$ acting on a set $A$ and meeting the following formula:

$$\forall x, y \in A \ [x \neq y \rightarrow R(x, y)] \text{ and } |A| \geq 2 \text{ (strong reciprocity)}$$

Further, they broaden their delimitation of reciprocity to include also those predicates which meet the required condition for some pair of $x$ and $y$ (rather than for all pairs of arguments), namely “if $x$ stands in relation $R$ to $y$, then $y$ also stands in relation $R$ to $x$”.

41
Figure 5. The mapping of semantic participants, valency complementations and surface positions in an example of syntactic reciprocity affecting ACT and ADDR of the verb nařknout 'to accuse':

(Petr a Pavel)\(\text{ACT}\) se \(\text{ADDR}\) vzájemně nařklí z podvodu.

‘Peter and Paul accused each other of deception.’

\approx\text{Petr} \text{ACT} nařkl Pavlo \text{ADDR} z podvodu a Pavel \text{ACT} nařkl Petra \text{ADDR} z podvodu.

‘Peter accused Paul of deception and Paul accused Peter of deception.’

**Syntactically reciprocal constructions**

In syntactically reciprocal constructions, one of the valency complementations involved in reciprocity is pluralized. This complementation is expressed either in the subject position (henceforth referred to as subject-oriented reciprocity), see examples (17-a) to (17-c), or in the direct object position (object-oriented reciprocity), see example (17-d). The other complementation affected by reciprocity – the one that is realized in a less prominent surface position – can be expressed by the reflexive personal pronoun or by the bipartite expression jeden – druhý ‘each other’, both marking the referential identity of this valency complementation and the one that is pluralized. As a result, the two valency complementations corefer.

---

21 However, let us remark that conventionalized reciprocal constructions expand to a broad range of situations, exceeding the one formally described in footnote 20. Namely, in the case where plural entities are involved in reciprocity, the relations between them can be configured in various ways. This issue has been largely discussed in formal semantics (see esp. Langendoen, 1978; Dalrymple et al., 1998). Different configurations of reciprocal relations are not, however, linguistically encoded, for Czech see esp. Panevová (2007) and for Slovak Ivanová (2020).

22 The bipartite expression jeden – druhý, which unambiguously marks reciprocity, is left aside here; see esp. Kettnerová and Lopatková (2020).
The reflexive pronoun, expressing the complementation in a less prominent surface position, can be either in the clitic or full form (compare examples (17-a) and (17-b)), depending on the morphological case and topic-focus articulation (see, e.g., Hajičová, 2007). However, in the case where this valency complementation has the comitative form, the less prominent surface position is mostly elided from the surface, see example (17-c).

(17)  
\[\text{a. V zajetí se šimpanzice navzájem brání proti násilí ze strany samců ...}
\]  
‘In captivity, female chimpanzees defend each other against male violence ...’  
\[\text{b. [... omývači mrtvol se naučili okrádat nejen nebožtíky,] nýbrž i sebe navzájem ...}
\]  
‘[... corpse washers have learned to rob not only the dead] but also each other ...’  
\[\text{c. Také v těžkých dobách spolu cítíme.}
\]  
‘We also sympathize with each other in difficult times.’  
\[\text{d. Spoje plosky nohou a dejte ruce k sobě.}
\]  
‘Join the soles of the feet and put your hands together.’

The syntactic operation transforming an underlying non-reciprocal construction into a reciprocal one is called \textit{syntactic reciprocalization}. As with syntactic reflexivization, syntactic reciprocalization is characterized by specific changes in the valency properties of verbs, namely changes in the morphosyntactic and lexical expression of their valency complementations (see Sections 5.1 and 5.2).

Syntactically reciprocal constructions with inherently reciprocal verbs

Besides conventionalized constructions, where the reciprocal situation is encoded by grammatical means, reciprocity can already be implied by the lexical meaning of some verbs, here called \textit{inherently reciprocal verbs} (see esp. Ševčíková, 2007; Paněvová and Mikulová, 2007). The term covers both verbs that are not derived by the word formation process of reflexivization and verbs that acquire the feature of reciprocity in this process. Inherently reciprocal verbs, containing the feature of reciprocity in their lexical meaning, express reciprocity for a particular pair (exceptionally two

\[\text{We thus distinguish between the reflexivization as a syntactic process producing syntactically reflexive constructions of a lexical unit of a verb (discussed in Section 4) and the reflexivization as a type of word formation process deriving a reflexive verb lemma from a non-reflexive one.}
\]

\[\text{Inherently reciprocal verbs that are not derived by the reflexivization of the latter type are referred to in the literature as symmetric predicates (König and Kokutani, 2006), lexical reciprocals (Nedjalkov, 2007) or allelic predicates (Haspelmath, 2007). Those that result from the word formation process of reflexivization and their verb lemmas are thus marked by the reflexives are subsumed under the so-called verb-marked reciprocals (Haspelmath, 2007) or morphological grammatical reciprocals (Nedjalkov, 2007).}
\]
pairs) of their valency complementations already in their basic constructions, where these valency complementations are expressed in separate syntactic positions. However, the reciprocal meaning of these constructions is merely implicated by inherently reciprocal verbs, and their reciprocal interpretation can be easily canceled, compare example (18-a) with a reciprocal reading and (18-b) without this interpretation. The reciprocal interpretation of these constructions is thus strongly contextually dependent (see Rákosi, 2008 and for Czech Panevová, 2007).

(18)  
  a. Gretchen právě hovořila s Richardem Sanfordem, ...
      ‘Gretchen was just talking with Richard Sanford, ...’
  b. Pachtinová s ním hovořila stísněně, jакši beз jískry, ...
      ‘Pachtin was talking with him uneasily, somehow without a spark, ...’
  c. Muži i ženy tam seděli u stolků nebo hovořili po skupinkách vstoje.
      ‘Men and women sat there at tables or talked standing up in groups.’

In Czech, inherently reciprocal verbs allow for syntactically reciprocal constructions resulting from the operation of syntactic reciprocalization as well. In contrast to reciprocal constructions of verbs without the feature of reciprocity in their lexical meaning, syntactically reciprocal constructions of inherently reciprocal verbs require less linguistic marking, see example (18-c), where only the pluralized valency complementation in the subject encodes reciprocity. As a result, separate rules must be formulated for the syntactic operation of reciprocalization covering inherently reciprocal verbs (see Section 5.2).

5.1. Syntactic Reciprocity in the Data Component

In the data component of VALLEX, the attribute recipro, assigned to individual lexical units of verbs, provides the information on syntactic reciprocity in the form of pairs of the affected valency complementations (2,744 lexical units in total, represented by 2,909 verb lemmas). As in the case of reflexivity (Section 4.1), the information on reciprocity is restricted to actants, quasi-actants and obligatory free modifications. Subject and object-oriented reciprocity are not explicitly distinguished here as this distinction follows from morphemic forms of the involved valency complementations. Similarly to reflexivity, subject-oriented reciprocity is much more frequent than object-oriented reciprocity, see Table 1 summarizing basic statistics on subject-oriented and object-oriented reciprocity contained in the VALLEX data.

---

Optional free modifications can be involved in reciprocity as well (e.g., Před ponořením si BEN navzájem kontrolovali vybavení. ‘Before the dive, they checked each other’s equipment.’ with the pro-dropped actant ACT and the optional free modification BEN involved in reciprocity). However, as optional free modifications are not systematically covered in the VALLEX lexicon, we leave them aside although it can be supposed that reciprocity involving optional free modification follows similar principles as actants and quasi-actants.
As in the case of reflexivity, reciprocity can affect more than one pair of valency complementations with a single lexical unit of a verb (from 2,744 lexical units of verbs with the attribute recipr, 222 can have more than one pair of valency complementations involved in reciprocity). In such cases, different pairs of valency complementations are captured in separate attributes recipr, distinguished by an Arabic numeral, see example (19). Further, a lexical unit can bear the feature of reciprocity in its lexical meaning for one pair but need not do so for the other. The information on the nature of the lexical unit with respect to this feature, captured by the value inherent in the attribute reciprverb, is thus relevant for individual pairs of the affected valency complementations (not for the whole lexical unit; this attribute is therefore attached to the respective pairs of valency complementations, see example (19-b).

(19) odpoutávat impf – odpoutat pf ‘to detach’ ... ACT1 PAT4 ORIG od-2

a. recipr1: ACT–ORIG
   ... přece jsme od sebe ORIG neodpoutali pohledy ...
   ‘... after all, we did not look away from each other ...’
   lit. ‘we did not detach from each other with our glances’

b. recipr2: PAT–ORIG, reciprverb2: inherent
   Během jednoho dne je třeba od sebe ORIG odpoutat bojující strany PAT.
   ‘Within one day, the warring parties must be detached from each other.’

<table>
<thead>
<tr>
<th>type</th>
<th>¬inherent pairs LUs</th>
<th>inherent pairs LUs</th>
<th>total pairs LUs</th>
</tr>
</thead>
<tbody>
<tr>
<td>subject</td>
<td>2,484 2,361</td>
<td>229 229</td>
<td>2,713 2,577</td>
</tr>
<tr>
<td>object</td>
<td>153 148</td>
<td>104 104</td>
<td>257 256</td>
</tr>
<tr>
<td>total</td>
<td>2,637 2,443</td>
<td>333 328</td>
<td>2,970 2,744</td>
</tr>
</tbody>
</table>

Table 1. Basic statistics on reciprocity (counted as pairs of the affected valency complementations and as the affected lexical units).

We can observe that, out of 222 lexical units with more than one pair of the valency complementations that can be involved in reciprocity, 27 lexical units represent inherently reciprocal verbs with respect to one pair of their valency complementations while they do not bear the feature of reciprocity for the other pair of complementations. Moreover, only five lexical units are characterized as inherently reciprocal verbs with respect to two distinct pairs of their complementations.
Subject-oriented reciprocity

In the case of subject-oriented reciprocity, the three pairs of valency complementations ACT–PAT, ACT–ADDR and ACT–DIR3 are attested in the VALLEX data most frequently with verbs that do not bear the feature of reciprocity in their lexical meaning, see examples (20-a) to (20-c). With inherently reciprocal verbs, the pairs of the actants ACT–PAT and ACT–ADDR are contained in the data most often, see examples (21-a) and (21-b).26

(20) a. bombardovat \textit{impf} ‘to bomb’ … ACT1 PAT4
    \begin{tabular}{l}
    \text{recipr: ACT–PAT} \\
    Obě válčící strany \text{ACT se PAT} navzájem bombardovaly.
    \end{tabular}
    \begin{tabular}{l}
    ‘Both warring parties bombed each other.’
    \end{tabular}
b. vyhrožovat \textit{impf} ‘to threat’ … ACT1 ADDR3 PAT7,2e,cont
    \begin{tabular}{l}
    \text{recipr: ACT–ADDR} \\
    \ldots, hráči \text{ACT si ADDR} vyhrožovali na střídačce i na ledě …
    \end{tabular}
    \begin{tabular}{l}
    ‘…, players threatened each other on the players’ bench and on the ice …’
    \end{tabular}
c. najíždět \textit{impf} – najet \textit{pf} ‘to run at’ … ACT1 DIR3
    \begin{tabular}{l}
    \text{recipr: ACT–DIR3} \\
    Zpěnění koně \text{ACT na sebe DIR} najíždějí v divokých skocích.
    \end{tabular}
    \begin{tabular}{l}
    ‘Frothy horses run at each other in wild jumps.’
    \end{tabular}

(21) a. ladit \textit{impf} ‘to fit’ … ACT1 PAT1,583
    \begin{tabular}{l}
    \text{recipr: ACT–PAT, recipverb: inherent} \\
    \ldots, aby (oděv a kabelka) \text{ACT spolu dokonale ladily}.
    \end{tabular}
    \begin{tabular}{l}
    ‘…, so that clothes and the handbag fit together perfectly.’
    \end{tabular}
b. soutěžit \textit{impf} ‘to compete’ … ACT1 ADDR proti3,7 PAT o+4
    \begin{tabular}{l}
    \text{recipr: ACT–ADDR, recipverb: inherent} \\
    \text{Přihlášení hráči PAT soutěžili ve dvojicích proti sobě ADDR}.
    \end{tabular}
    \begin{tabular}{l}
    ‘The registered players competed in pairs against each other.’
    \end{tabular}
\textit{Přihlášení hráči PAT soutěžili ve dvojicích.} (modified)
\begin{tabular}{l}
    ‘The registered players competed in pairs.’
    \end{tabular}

26 The following table gives numbers of pairs of the valency complementations affected by subject-oriented reciprocity, as attested in VALLEX – on the left are the pairs for which their governing lexical units do not bear the feature of reciprocity in their lexical meaning (with respect to the given pair, \textit{¬}inherent), on the right the pairs for which their governing lexical units contain this feature (inherent):

\begin{tabular}{|c|c|c|c|c|}
\hline
\text{–inherent} & \text{inherent} & \\
\hline
ACT–PAT & 1,583 & ACT–DIR3 & 144 & ACT–PAT & 117 & ACT–DIR3 & 2 \\
ACT–ORIG & 80 & ACT–DIR1 & 34 & \\
ACT–EFF & 24 & ACT–DIR2 & 7 & \\
ACT–OBST & 15 & \\
\hline
\end{tabular} 
In syntactically reciprocal constructions, the pluralized valency complementation, typically \( ACT \), is expressed as the (nominative) subject. The other valency complementation (i.e., the one that is realized in a less prominent surface position) is expressed depending on the lexical meaning of its governing lexical unit of the verb: (i) In the case of the lexical unit without the feature of reciprocity for the involved pair of complementations, the other valency complementation is expressed on the surface by the reflexive personal pronoun in the morphemic form determined for this complementation in the valency frame of the verb (the clitic or full form of the reflexive depends on the morphological case and on the topic-focus articulation, cf. Hajičová, 2007), see examples (20-a) to (20-c). In limited cases, the role of the reciprocal marker is taken over by adverbials expressing reciprocity.\(^{27}\) (ii) If the lexical unit bears the feature of reciprocity of the affected valency complementations in its lexical meaning, the other valency complementation is only optionally expressed on the surface by the reflexive pronoun, compare examples in (21-b). In particular, when this complementation has the comitative form \( s+7 \), it is predominantly deleted from the surface, see example (21-a). Syntactically reciprocal constructions of inherently reciprocal verbs thus require less linguistic marking, see Section 5.2 below.

As to the surface position of the other valency complementation, it can be expressed as the direct object in the accusative, see example (20-a), as an indirect object, see examples (20-b) and the first example in (21-b), or as an adverbial, see example (20-c). The following morphemic forms are attested in the data for those actants and quasi-actants of lexical units that do not bear the feature of reciprocity in their lexical meaning: 2, 3, 4, 7, bez+2, do+2, k+3, kolem+2, mezi+4, na+4, na+6, nad+4, nad+7, o+4, o+6, od+2, po+6, podle+2, pro+4, proti+3, před+4, před+7, při+6, s+7, u+2, v+4, v+6, z+2, za+4, za+7 and the following ones for actants and quasi-actants of lexical units that bear the feature of reciprocity: 3, 7, k+3, od+2, proti+3, s+7, z+2.\(^{28, 29}\)

\(^{27}\) Adverbials as the primary marker of reciprocity occur in two cases. First, they can express reciprocity in syntactically reciprocal constructions of verbs with reflexive lemmas where the clitic reflexive is subject to haplology (compare, e.g., A pak jsme si sobě postěžovali, ... and A pak jsme si navzájem postěžovali, ... 'And then we complained to each other'). Second, the comitative, i.e., the form \( s+7 \) in Czech, differs in this respect as well: either the complementation in the comitative or an adverbial expressing reciprocity is present on the surface as the primary marker of reciprocity (compare, e.g., ..., než by se sebou nesouhlasili, ... '..., rather than disagree with each other, ...' and V mnoha věcech jsme spolu nesouhlasili, ... 'We disagreed with each other on many things, ...'

\(^{28}\) Morphemic forms of obligatory free modifications are not explicitly specified.

\(^{29}\) As for the frequency of individual morphemic forms, prepositionless cases are attested with 1,571 LUs (1,012 accusatives, 526 datives, 52 genitives, and 25 instrumentals; note that more than one of these prepositionless cases can express a single affected valency complementation). Further, the listed prepositional cases occur with 960 LUs (the most frequent being \( s+7 \) (288 LUs), \( na+4 \) (223 LUs), and \( o+6 \) (116 LUs)).
Object-oriented reciprocity

In the case of object-oriented reciprocity, the three pairs of valency complementations illustrated in examples (22-a) to (22-c) are attested in the VALLEX data most frequently with verbs that do not bear the feature of reciprocity in their lexical meaning. Examples (23-a) to (23-c) then illustrate the pairs of valency complementations most frequently affected by reciprocity with verbs that bear the feature of reciprocity (with respect to these pairs):  

30

(22)  
a. \(dávat\) impf – \(dát\) pf ‘to put’ ... \(\text{ACT}_1\) PAT\(_4\) DIR3  
   recip: PAT–DIR3  
   \(\text{Spojte plosky nohou a dejte ruce pat k sobě dir3}.\) (= (17-d))  
   Join the soles of the feet and put your hands together.

b. \(přizpůsobovat\) impf – \(přizpůsobit\) pf ‘to adjust’ ... \(\text{ACT}_1\) PAT\(_4\) EFF\(_3\)  
   recip: PAT–EFF  
   \(\text{Je jasné, že musí být skutečně velmi složité přizpůsobit navzájem dva tak samostatné (učební a vyučující) nástroje pat, ...}\)  
   ‘It is clear that it must be really very difficult to adapt two so separate (learning and teaching) tools to each other, ...’

c. \(odstěhovat\) impf ‘to move away’ ... \(\text{ACT}_1\) PAT\(_4\) DIR1  
   recip: PAT–DIR1  
   \(\text{Někdy je nutné rozhádané klienty pat od sebe dir1 odstěhovat.}\)  
   ‘Sometimes it is necessary to move quarrelsome clients apart.’

(23)  
a. \(sbližovat\) impf – \(sblížit\) pf ‘to bring closer’ ... \(\text{ACT}_1\) ADDR\(_{s+7}\) PAT\(_4\)  
   recip: ADDR–PAT, recipverb: inherent  
   (Moskva a Peking) pat \(\text{ale incident sbližil ...}\)  
   ‘The incident, however, brought Moscow and Beijing closer ...’

b. \(ztotožňovat\) impf – \(ztotožnit\) pf ‘to identify’ ... \(\text{ACT}_1\) PAT\(_4\) EFF\(_{s+7}\)  
   recip: PAT–EFF, recipverb: inherent  
   \(\text{Není možné ztotožňovat (islám a terorismus) pat.}\)  
   ‘It is not possible to identify Islam and terrorism.’

c. \(oddělovat\) impf – \(oddělit\) pf ‘to separate’ ... \(\text{ACT}_1\) PAT\(_4\) ORIG\(_{od-2}\)  
   recip: PAT–ORIG, recipverb: inherent

The following table gives the number of pairs of the valency complementations affected by object-oriented reciprocity, as attested in VALLEX – the left part displays those pairs for which their governing lexical units do not bear reciprocity in their lexical meaning, the right part the ones for which they contain this feature:

<table>
<thead>
<tr>
<th></th>
<th>inherent</th>
<th>inherent</th>
<th>inherent</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAT–EFF</td>
<td>20</td>
<td>PAT–DIR3</td>
<td>79</td>
</tr>
<tr>
<td>ADDR–PAT</td>
<td>8</td>
<td>PAT–DIR1</td>
<td>13</td>
</tr>
<tr>
<td>PAT–ORIG</td>
<td>1</td>
<td>PAT–LOC</td>
<td>12</td>
</tr>
<tr>
<td>PAT–OBST</td>
<td>9</td>
<td>PAT–DIR2</td>
<td>3</td>
</tr>
<tr>
<td>ADDR–PAT</td>
<td>53</td>
<td>PAT–DIR3</td>
<td>4</td>
</tr>
<tr>
<td>PAT–EFF</td>
<td>27</td>
<td>PAT–DIR1</td>
<td>4</td>
</tr>
<tr>
<td>PAT–ORIG</td>
<td>15</td>
<td>PAT–LOC</td>
<td>1</td>
</tr>
<tr>
<td>PAT–OBST</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
V. Kettnerová, M. Lopatková, A. Vernerová  
Reflexives in VALLEX (27–60)

[... počkal si] na vlnu, která na chvíli oddělila obě lodě PAT od sebe ORIG, ...

‘[... he waited] for the wave that separated the two ships from each other for a moment, ...’

In syntactically reciprocal constructions, the pluralized valency complementation is mostly expressed as the (accusative) direct object (22-a) to (22-c) and (23-a) to (23-c). Further, it can be sporadically realized as an indirect object in the dative (24-a), in the instrumental (the first example in (24-b)), or in the prepositional group s+7 (the second example in (24-b)). The other valency complementation, occupied by the reflexive personal pronoun, is realized on the surface depending on the lexical meaning of the verb and on the morphemic forms of this complementation. First, if the verb does not bear the feature of reciprocity in its lexical meaning, this valency complementation is realized as an indirect object in one of the following forms: 3, do+2, k+3, na+4, na+6, o+4, o+6, od+2, po+6, proti+3, s+7, před+7, v+4, see examples (24-a) and (24-b), or as an adverbial, see examples (22-a) and (22-c). Subjectless constructions are sporadically attested with object-oriented reciprocity as well, see example (24-a). Second, with inherently reciprocal verbs, this complementation mostly remains unexpressed (inherently reciprocal verbs thus require less linguistic marking, similarly to subject-oriented reciprocity). If it is realized on the surface, it is expressed as an indirect object with the following forms: 3, k+3, od+2 and proti+3, see example (23-c), or as an adverbial, see example (24-c).33

(24)  

a. záležet impf ‘to care’ ... ACT 3 PAT na+6, aby  
recipr: ACT–PAT  
Toužím po pěkném vztahu dvou lidí, kterým ACT na sobě PAT opravdu záleží.  
‘I long for a nice relationship between two people who really care about each other.’

b. mlátit impf ‘to beat’ ... ACT 3 PAT 7, s+7 OBST o+4  
recipr: PAT–OBST  
... bral do rukou dva kameny a mlátil jimí PAT o sebe OBST, ...

31 The only exception represents cases in which this valency complementation has the dative form. In this case, either the full form of the reflexive pronoun in the dative or one of the adverbials expressing reciprocity serves as the primary marker of reciprocity, see Table 3 in Section 5.2.

32 As in the case of reflexivity, this inventory of morphemic forms comprises only the forms of actants and quasi-actants.  
As for the frequency of individual morphemic forms, the dative is attested with 6 LUs and prepositional cases with 44 LUs.

33 In the VALLEX data, the dative is attested with 4 LUs and the prepositional cases k+3, od+2, proti+3 with 33 LUs of inherently reciprocal verbs.  
In addition, in non-reciprocal constructions, the other valency complementation can be expressed by the morphemic forms 7, s+7 and za+4; however, the first two are obligatorily deleted from the surface in reciprocal constructions and the third one, s+7 (listed in 62 LUs), can only sporadically be expressed (see also Table 3, Section 5.2).
‘..., he took two stones in his hands and smashed them against each other, ...

[Pořád si ode mě půjčuje ty medaile] a dost s nimi pat o sebe obst mlátí.
‘[He keeps borrowing those medals from me] and banging them against each other a lot.’

c. přibližovat impf – přibližit pf ‘to bring closer’ ... ... ACT 1 PAT 4 DIR3
recipr: PAT–DIR3, reciprverb: inherent
Sociální sítě přibližily lidi pat k sobě DIR3 tak blízko, ...
‘Social networks brought people so close, ....’

5.2. Syntactic Reciprocity in the Grammar Component

Syntactic reciprocity in the grammar component of VALLEX is described by two basic rules – one for subject-oriented syntactic reciprocity and the other for object-oriented reciprocity – applied to the valency frames of verbs underlying their non-reciprocal constructions. These basic rules describe changes concerning the valency complementation that is realized in the more prominent surface position involved in reciprocity. Both these basic rules are complemented by a set of supplementary rules, applied to the valency frames output by the basic rules. The supplementary rules determine changes related to the valency complementation expressed in the less prominent position.

Each rule consists of three sections (see Figures 6, 7 and Tables 2, 3):

A header identifies the rule.

Conditions indicate properties that a lexical unit of a verb and its valency frame must have so that the rule can be applied. Four properties are relevant here:

- First, the lexical unit must be characterized by the attribute recipr identifying a pair of the valency complementations affected by reciprocity; these complementations are represented by variables X and Y in the case of subject-oriented reciprocity (with X reserved for the nominative complementation), and by Y and Z in the case of object-oriented reciprocity.
- Second, in both types of rules, valency complementations are restricted by morphemic forms, introduced in their subscripts.
- Third, in the supplementary rules, it is specified whether these rules can be applied to verbs without the feature of reciprocity (¬reciprverb: inherent), or to inherently reciprocal verbs (reciprverb: inherent).
- Fourth, in the supplementary rules for subject-oriented reciprocity only, conditions are imposed on verb lemmas, namely whether the rules are applicable to lexical units with reflexive lemmas (the abbreviation SE|SI), or to units with non-reflexive lemmas (¬SE|SI).

A set of actions indicates changes in the valency frames of verbs underlying their non-reciprocal constructions necessary to generate the frames underlying their syntactically reciprocal constructions.
Three types of changes are relevant here:

- First, the valency complementation X in the case of subject-oriented reciprocity and Y in the case of object-oriented reciprocity is pluralized by the respective basic rules.\(^{34}\)
- Second, the supplementary rules stipulate that the valency complementation Y in the case of subject-oriented reciprocity and Z in the case of object-oriented reciprocity are filled by the reflexive personal pronoun in the respective morphemic form prescribed for this complementation; the abbreviation REF, specified in the subscript by the index prep-case, is used for the form of the reflexives in prepositional groups, except for the comitative s+7 (this form is treated by special rules). Alternatively, these valency complementations have the null form, i.e., they are not expressed on the surface (marked as \(\emptyset\)). Finally, other possible morphemic forms of all the valency complementations X, Y and Z involved in reciprocity, if available, are deleted.
- Third, the basic and supplementary rules determine that all the affected valency complementations (X and Y for subject-oriented reciprocity, or Y and Z for object-oriented reciprocity) must be obligatorily present in the deep structure of syntactically reciprocal constructions (the keyword “obligatoriness”).
- In addition, the supplementary rules indicate a list of adverbials emphasizing a reciprocal interpretation (the keyword “emphasis”).\(^{35}\)

If the valency complementation Y in the case of subject-oriented reciprocity and Z in the case of object-oriented reciprocity can be expressed by more than one prepositionless or prepositional case, all possible pairs of the respective basic rule and the respective supplementary rule are applied, generating more than one valency frame underlying reciprocal constructions.

Rules for subject-oriented reciprocity

The basic rule rcpr_sbj_basic, given in Fig. 6, describes changes concerning the valency complementation X in reciprocal constructions. Further, the supplementary rules defined for subject-oriented reciprocity are summarized in Table 2; the table presents the means of expressing reciprocity depending on the morphemic form of the valency complementation Y, on the lemma of its governing verb (whether it is reflexive or not; the symbol “—” stands for cases where this distinction is not relevant), and on the type of the verb (inherently reciprocal or not). The fourth column

\(^{34}\) As X, reserved for the nominative complementation, functions as subject, the basic rule for subject-oriented reciprocity must cover also the change in verb agreement. For the sake of simplicity, we leave this issue aside.

\(^{35}\) For the sake of simplicity, we omit this information from summarizing Tables 2 and 3.
of the table gives the forms of the reflexive personal pronoun and adverbials expressing reciprocity (∅ indicates that reciprocity, being expressed by the pluralized X, does not require additional marking). Fully formalized rules can be found in Lopatková et al. (2021).

The basic rule rcpr_basic_subj is applied to all cases of subject-oriented reciprocity recorded in the VALLEX data, namely it concerns 2,713 pairs of valency complementations in 2,577 lexical units. Table 2 then provides numbers of pairs of valency complementations and numbers of lexical units covered by the individual supplementary rules.

Subject-oriented reciprocity

rcpr_sbj_basic

<table>
<thead>
<tr>
<th>recipr:</th>
<th>X–Y</th>
<th>X₁ &amp; Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>*→ 1: plural</td>
<td>X</td>
</tr>
<tr>
<td>obligatoriness</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 6. The basic rule rcpr_sbj_basic indicating changes in valency frames of verbs in the case of subject-oriented reciprocity.

<table>
<thead>
<tr>
<th>input</th>
<th>reflexive verb</th>
<th>inherently reciprocal</th>
<th>expression of reciprocity</th>
<th># pairs</th>
<th># LUs</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>N</td>
<td>N</td>
<td>si, sobě</td>
<td>462</td>
<td>462</td>
</tr>
<tr>
<td>3</td>
<td>Y</td>
<td>N</td>
<td>sobě, navzájem, vzájemně, mezi sebou</td>
<td>57</td>
<td>57</td>
</tr>
<tr>
<td>3</td>
<td>Y</td>
<td>Y</td>
<td>sobě, ∅</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>—</td>
<td>N</td>
<td>se, sebe</td>
<td>1,012</td>
<td>1,012</td>
</tr>
<tr>
<td>2</td>
<td>—</td>
<td>N</td>
<td>sebe</td>
<td>52</td>
<td>52</td>
</tr>
<tr>
<td>7</td>
<td>—</td>
<td>N,Y</td>
<td>sebou</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>s+7</td>
<td>—</td>
<td>N</td>
<td>se sebou, navzájem, vzájemně, mezi sebou, spolu</td>
<td>75</td>
<td>75</td>
</tr>
<tr>
<td>s+7</td>
<td>—</td>
<td>Y</td>
<td>∅, se sebou</td>
<td>213</td>
<td>213</td>
</tr>
<tr>
<td>prep+case</td>
<td>—</td>
<td>N</td>
<td>REFL prep-case</td>
<td>860</td>
<td>665</td>
</tr>
<tr>
<td>prep+case</td>
<td>—</td>
<td>Y</td>
<td>REFL prep-case, ∅</td>
<td>20</td>
<td>20</td>
</tr>
</tbody>
</table>

Table 2. The supplementary rules for subject-oriented reciprocity.

Let us illustrate how the valency frame underlying syntactically reciprocal constructions of the verb rovnat se^impf ‘to equal’ is derived. For the lexical unit of the verb
Reflexives in VALLEX (27–60)

rovnat se 

‘to equal’ with the valency frame in (25), it is determined in the data component of the lexicon that two of its actants, ACT and PAT, can be involved in reciprocity. As ACT is in the nominative, the basic rule for subject-oriented reciprocity rcrpr_sbj_basic rule is applied. The rule identifies ACT as X and stipulates that this nominative ACT is pluralized and must be present in the deep structure of syntactically reciprocal constructions (but can be deleted from the surface); see Fig. 6. Subsequently, the supplementary rule covering the dative Y for inherently reciprocal verbs with reflexive lemmas is applied, selected on the basis of the following three properties of the lexical unit (see Table 2). First, PAT is in the dative. Second, reciprocity of ACT and PAT is marked by the attribute recipverb with the value inherent as encoded in the lexical meaning of the lexical unit. Third, the verb lemma representing this lexical unit contains the reflexive verbal affix se. This supplementary rule then states that PAT must be present in the deep structure of reciprocal constructions and is occupied by the full form of the reflexive personal pronoun sobě or has the null form, thus remaining unexpressed on the surface, see the resulting valency frame (26), underlying the reciprocal constructions in (26-a) and in (26-b).

(25) rovnat se 

‘to equal’ ...

recip: ACT–PAT, recipverb: inherent

... rozměr ACT strany základny jehlanu se rovná průměru PAT základny kužele.

‘... the length of the side of the base of the pyramid is equal to the diameter of the base of the cone.’

(26) ACT1: plural PATsobě,∅

a. ... zda se sobě PAT strany ACT rovnice rovnaži.

‘... whether the sides of the equation are equal to each other.’

b. V praxi se však tyto podíly ACT nerovnaji.

‘In practice, however, these proportions are not equal to each other.’

Rules for object-oriented reciprocity

The basic rule rcrpr_obj_basic, given in Fig. 7, characterizes changes of the valency complementation Y in the case of object-oriented reciprocity. The basic rule is complemented by six supplementary rules summarized in Table 3. The supplementary rules present the means of expressing reciprocity depending on the morphemic form of the valency complementation Z (column headed with Z), and on the type of its governing verb (inherently reciprocal or not). The third column gives the forms of the reflexive personal pronoun and adverbials expressing reciprocity (∅ indicates

36 In the case of the null form of PAT, the rule copes with the fact that the clitic form of the reflexive pronoun in the dative sí (in contrast to its full form sobě) is subject to haplology, i.e., it fuses with the reflexive verbal affix se as part of the verb lemma (see Rosen, 2014). As a result, the resulting reciprocal construction contains only a single reflexive se, see example (26-b).
that reciprocity is expressed by the pluralized $Y$ only and does not require further marking). Fully formalized rules can be found in Lopatková et al. (2021).

<table>
<thead>
<tr>
<th>Reciprocity $Y-Z$</th>
<th>rcpr_obj_basic</th>
</tr>
</thead>
<tbody>
<tr>
<td>recipr:</td>
<td>$Y-Z$</td>
</tr>
<tr>
<td></td>
<td>$Y_{3,4,7,s+7}$</td>
</tr>
<tr>
<td>$Y$</td>
<td>$3 \rightarrow 3$: plural</td>
</tr>
<tr>
<td></td>
<td>$4 \rightarrow 4$: plural</td>
</tr>
<tr>
<td></td>
<td>$7 \rightarrow 7$: plural</td>
</tr>
<tr>
<td></td>
<td>$s+7 \rightarrow s+7$: plural</td>
</tr>
<tr>
<td>obligatoriness</td>
<td>$Y$</td>
</tr>
</tbody>
</table>

Figure 7. The basic rule rcpr_obj_basic indicating changes in valency frames of verbs in the case of object-oriented reciprocity.

The basic rule rcpr_basic_obj is applied to all cases of object-oriented reciprocity recorded in the VALLEX data, namely it concerns 257 pairs of valency complementations in 256 lexical units. Table 3 then provides numbers of pairs of valency complementations and numbers of lexical units covered by the individual supplementary rules.

Let us illustrate the rule for object-oriented reciprocity with one of the lexical units of the verb vystřídat$^{pf}$. This lexical unit is characterized by the reciprocity of PAT and EFF, see the valency frame of this unit (27), underlying its non-reciprocal constructions. To generate the valency frame describing its reciprocal constructions, the basic rule for object-oriented reciprocity, the rule rcpr_obj_basic, is used (see Fig. 7). This rule identifies PAT as $Y$, changes its number into plural and stipulates that it is obligatorily present in both deep and surface structure of reciprocal constructions. Further, this basic rule is supplemented with the rule determined for $Z$ in the instrumental with inherently reciprocal verbs, as given in Table 3. According to this rule, EFF in the instrumental has the null form, i.e., it remains unexpressed on the surface, see the resulting valency frame in (28) and the example of a reciprocal construction given there.

(27) vystřídat$^{pf}$ ‘to change’ … ACT$_1$ PAT$_4$ EFF$_{7,s+7,2a+4}$

recipr: PAT–EFF, recipverb: inherent

..., když [trenér Brückner] vystřídal záložníka Šmicer.PAT obráncem EFF Jiránekem, ...

‘...when [coach Brückner] replaced midfielder Šmicer with defender Jiránek, ...

...’
V. Kettnerová, M. Lopatková, A. Vernerová

Reflexives in VALLEX (27–60)

<table>
<thead>
<tr>
<th>input form of Z</th>
<th>inherently reciprocal</th>
<th>expression of reciprocity</th>
<th># pairs</th>
<th># LUs</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Y</td>
<td>Ø, sobě</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>N</td>
<td>sobě or one of the adverbials navzájem, vzájemně</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>Y</td>
<td>Ø</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>s+7</td>
<td>N,Y</td>
<td>Ø, se sebou</td>
<td>63</td>
<td>63</td>
</tr>
<tr>
<td>prep+case</td>
<td>Y</td>
<td>REFL prep-case and/or †† Ø</td>
<td>43</td>
<td>43</td>
</tr>
<tr>
<td>prep+case</td>
<td>N</td>
<td>REFL prep-case</td>
<td>46</td>
<td>46</td>
</tr>
</tbody>
</table>

†† The full rule specifies the choice of a morphemic form (REFL prep-case or the null form) depending on the input prepositional case (see Lopatková et al., 2021 and footnote 33).

Table 3. The supplementary rules for object-oriented reciprocity.

(28) \( \text{ACT}_1 \text{ PAT}_4 \text{ plural EFF} (\emptyset) \)

‘Vystřídala jsem už tři různé zaměstnavatele. PAT.’

As the actant EFF has the form of the prepositional groups s+7 and za+4 as well, the supplementary rules for these forms are applied to the output of the basic rule, too: the supplementary rule for the form za+4 gives the same output valency frame as the rule for the instrumental case already shown in (28) while the supplementary rule for the form s+7 allows (in addition to the null form) the reflexive personal pronoun se sebou to be expressed on the surface (however, this form is very rare).

6. Reflexive-reciprocal Ambiguity

As the reflexive personal pronoun in Czech encodes both reflexivity and reciprocity, syntactic constructions marked by the reflexives can – under certain conditions – be ambiguous, i.e., they can have either a reflexive or a reciprocal reading. Ambiguous constructions are formed by lexical units of verbs for which the same pair of valency complementations can be involved in both reflexivity and reciprocity. Then if the valency complementation expressed in the more prominent syntactic position is plural and at the same time the one in the less prominent position is expressed by the reflexive personal pronoun, the resulting construction is either reflexive or reciprocal, depending on the context.

For example, with the lexical unit of the verb odpoušťet \textit{impf} – odpustit \textit{pf}, ACT and ADDR can be affected by both reflexivity and reciprocity, see the valency frame and the reflex and recipr attributes in (29). If ACT is plural and ADDR is expressed by the reflexive pronoun, the resulting construction has two interpretations: it is either re-

37 See also footnote 33.
flexive or reciprocal, see example (29-a). To disambiguate its meaning, the following expressions are used: the pronoun sám ‘oneself’ for the reflexive reading and the adverbials vzájemně, navzájem, spolu or mezi sebou for the reciprocal reading, compare examples (29-b) and (29-c).

(29) odpouštět impf – odpustit pf ‘to forgive’ … ACT₁ ADDR₃ PAT₄,₂e

reflex: ACT–ADDR
recipr: ACT–ADDR

a. Odpusťme si ADDR naše viny …
   ‘Let us forgive ourselves our guilt …’ or
   ‘Let us forgive each other our guilt …’

b. Odpusťme sami sobě ADDR, … (modified)
   ‘Let us forgive ourselves, …’

c. Odpusťme si ADDR navzájem … (modified)
   ‘Let us forgive each other …’

In the VALLEX data, 1,768 lexical units of verbs make it possible to create both reflexive and reciprocal constructions. Out of these, 1,653 lexical units are characterized by reflexivity and reciprocity affecting the same pair of valency complementations (1,738 pairs in total). These lexical units thus form ambiguous constructions with either a reflexive or a reciprocal interpretation. The rules contained in the grammar component of the lexicon then determine the linguistic means disambiguating their reading (see Lopatková et al., 2021).

7. Conclusion

This paper has provided a thorough description of the lexicographic representation of two valency changing phenomena, reflexivity and reciprocity, in VALLEX, the valency lexicon of Czech verbs. This representation makes use of two components of the lexicon: the data and the grammar component.

In the data component, around one third of the lexical units of verbs are assigned the attributes reflex and/or recipr, providing the information on pairs of valency complementations that can be involved in reflexivity and/or reciprocity (namely, 2,039 lexical units with reflex and 2,744 with recipr out of 6,829 lexical units in total). Each lexical unit can be assigned more than one attribute reflex and more than one attribute recipr, reflecting the fact that lexical units allow different pairs of their valency complementations to be affected by these phenomena. If a lexical unit bears the feature

38 The following pairs of valency complementations can be affected by both reflexivity and reciprocity:

<table>
<thead>
<tr>
<th>ACT–PAT</th>
<th>1,208 LUs</th>
<th>PAT–EFF</th>
<th>6 LUs</th>
<th>ACT–DIR₁</th>
<th>36 LUs</th>
<th>PAT–DIR₁</th>
<th>2 LUs</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACT–ADDR</td>
<td>425 LUs</td>
<td>ADDR–PAT</td>
<td>1 LU</td>
<td>ACT–DIR₃</td>
<td>8 LUs</td>
<td>ACT–LOC</td>
<td>8 LUs</td>
</tr>
<tr>
<td>ACT–ORIG</td>
<td>27 LUs</td>
<td></td>
<td></td>
<td>ACT–DIR₁</td>
<td>6 LUs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACT–EFF</td>
<td>19 LUs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

56
of reciprocity for some of its valency complementations in its lexical meaning, the attribute recipr is coupled with the attribute reciprverb with the value inherent, indicating that the lexical unit behaves with respect to this pair of valency complementations as an inherently reciprocal verb (328 lexical units).

The grammar component then stores rules that allow the valency frames underlying reflexive and reciprocal constructions of these lexical units of verbs to be derived from the valency frames describing their non-reflexive and non-reciprocal constructions (in the case of inherently reciprocal verbs, they reflect that these verbs require less linguistic marking).

As Czech uses reflexives for expressing both reflexivity and reciprocity, some lexical units can create ambiguous constructions that have either a reflexive or a reciprocal reading. These lexical units (1,653 in total) have the same pairs of valency complementations recorded in the attributes reflex and recipr. The rules then provide a list of the linguistic means disambiguating between a reflexive and a reciprocal interpretation.

Acknowledgements

We would like to thank the reviewers of the paper for their helpful comments. Markéta Lopatková’s contribution to work described herein has been supported by the grant No. GX20-16819X of the Czech Science Foundation (GAČR) Language Understanding: From Syntax to Discourse (LUSyD). The work on this project has also been supported by the Ministry of Education, Youth and Sports of the Czech Republic, Project No. LM2018101 LINDAT/CLARIAH-CZ. The work on the VALLEX lexicon has been using data and tools provided by the LINDAT/CLARIAH-CZ Research Infrastructure (https://lindat.cz), supported by the Ministry of Education, Youth and Sports of the Czech Republic (Project No. LM2018101).

Bibliography


58


Address for correspondence:
Václava Kettnerová
kettnerova@ufal.mff.cuni.cz
Institute of Formal and Applied Linguistics
Matematicko-fyzikální fakulta
Malostranské náměstí 25
118 00 Praha 1
Czech Republic
INSTRUCTIONS FOR AUTHORS

Manuscripts are welcome provided that they have not yet been published elsewhere and that they bring some interesting and new insights contributing to the broad field of computational linguistics in any of its aspects, or of linguistic theory. The submitted articles may be:

- long articles with completed, wide-impact research results both theoretical and practical, and/or new formalisms for linguistic analysis and their implementation and application on linguistic data sets, or
- short or long articles that are abstracts or extracts of Master’s and PhD thesis, with the most interesting and/or promising results described. Also
- short or long articles looking forward that base their views on proper and deep analysis of the current situation in various subjects within the field are invited, as well as
- short articles about current advanced research of both theoretical and applied nature, with very specific (and perhaps narrow, but well-defined) target goal in all areas of language and speech processing, to give the opportunity to junior researchers to publish as soon as possible;
- short articles that contain contraversing, polemic or otherwise unusual views, supported by some experimental evidence but not necessarily evaluated in the usual sense are also welcome.

The recommended length of long article is 12–30 pages and of short paper is 6–15 pages.

The copyright of papers accepted for publication remains with the author. The editors reserve the right to make editorial revisions but these revisions and changes have to be approved by the author(s). Book reviews and short book notices are also appreciated.

The manuscripts are reviewed by 2 independent reviewers, at least one of them being a member of the international Editorial Board.

Authors receive a printed copy of the relevant issue of the PBML together with the original pdf files.

The guidelines for the technical shape of the contributions are found on the website http://ufal.mff.cuni.cz/pbml. If there are any technical problems, please contact the editorial staff at pbml@ufal.mff.cuni.cz.