Abstract: Natural languages are typically characterized by a large area where grammar and lexicon overlap. Complex predicates with light verbs represent a typical language phenomenon at the lexicon-grammar interface. Their theoretically adequate representation thus requires a close interplay between the lexicon and the grammar. In this paper, we introduce a formal model for the lexicographic description of Czech complex predicates of the given type. The central type of Czech complex predicates are composed of light verbs and predicative nouns. We demonstrate that although their syntactic structure formation is highly complex, it still exhibits enough regularity to be captured by formal rules.

1 Motivation

Complex predicates with light verbs (CPs) consist of two syntactic units, a light verb (LV) and a predicative noun (PN) (or, sporadically, a predicative adjective or adverb); this verb-noun pair forms a single predicative unit, as for example dát radu ‘give advice’, dostat rozkaz ‘get an order’, mít radost ‘be happy’ (lit. have joy), or uzavřít dohodu ‘make an agreement’. Due to their complex characteristics, CPs proven to be challenging for syntactic theories as well as for natural language processing tasks.

Complex predicates with light verbs are characterized by a discrepancy in their syntax and semantics [1]: whereas the meaning of a CP is primarily expressed by the predicative noun, forming thus the semantic core of the CP, it is the semantically impoverished light verb which serves as the syntactic center of a sentence. We can exemplify this discrepancy on the CP uzavřít dohodu ‘make an agreement’, as used in (1). This CP is semantically characterized by three participants, namely ‘Party_1’ (Francie ‘France’), ‘Party_2’ (Německo ‘Germany’), and ‘Obligation’ (neútočení ‘non-aggression’), all these participations to the syntactic structure formation of CPs since morphological forms of valency complementations serve as valuable clues for determining whether a certain valency complementation belongs to the light verb or to the predicative noun. However, none of the works focused on Czech CPs provides an explicit description of the syntactic structure formation of CPs, see esp. [6, 7].

Although the contribution of light verbs and predicative nouns to the syntactic structure formation of CPs has been put under scrutiny within various theoretical frameworks – see e.g. argument merger formulated within the Government Binding theory [2], argument fusion [3] and argument composition within the Lexical-Functional Grammar [4], and the study by Alonso Ramos drawing on the Meaning ↔ Text Theory [5] – many of its aspects still remain unclear.

Czech, as an inflectional language encoding syntactic relations by morphological cases, provides a great opportunity to study the distribution of valency complementations in syntactic structures of CPs since morphological forms of valency complementations serve as valuable clues for determining whether a certain valency complementation belongs to the light verb or to the predicative noun. However, none of the works focused on Czech CPs provides an explicit description of the syntactic structure formation of CPs, see esp. [6, 7].

In this paper, we summarize our theoretical results described earlier and relate the proposed model with an extensive data annotation, see esp. [8, 9, 10]. We focus on the deep and surface structure of CPs, mainly with respect to the contribution of valency complementations to the syntactic structure of CPs made by the light verb and by the predicative noun and with respect to the role of coreference between the complementations in these structures (Section 3). On the basis of our theoretical findings, we propose an economic and linguistically informed formal model of CPs consisting of a grammatical part (Section 3) and
a lexical part (Section 4). Finally, grounded on extensive data annotation, we introduce an overall typology of CPs based on their coreferential characteristics and provide basic statistics for Czech CPs (Section 5).

2 VALLEX and FGD Framework

The proposed representation of CPs is elaborated within the Functional Generative Description (FGD), a stratificational and dependency-oriented theoretical linguistic framework [11]. One of the core concepts of FGD is that of valency [12]: at the layer of linguistically structured meaning (also the deep syntactic layer), it is the valency that provides the structure of a dependency tree. The valency structure of FGD has been applied in several valency lexicons, esp. PDT-Vallex1 [13] and VALLEX2 [14], and verified on extensive corpus data, esp. within the Prague Dependency Treebank (PDT)3. VALLEX, being the most elaborated lexicon of Czech verbs, forms a solid basis for the lexical component of FGD.

For the purpose of representation of language phenomena bridging between the grammar and the lexicon (e.g., diatheses and reciprocity), VALLEX is divided into a lexical part (i.e., the data component) and a grammatical part (i.e., the grammar component) [15, 16]. This division proves to be useful also for the representation of CPs.

Data component. The central organizing concept of the lexical part of VALLEX is the concept of lexeme. The lexeme associates a set of lexical forms, representing the verb in an utterance, with a set of lexical units, corresponding to their individual senses.

The data component consists of an inventory of lexical units of verbs with their respective valency frames underlying their deep syntactic structures. Each valency frame is modeled as a sequence of frame slots corresponding to valency complementations of a verb labeled by (rather coarse-grained) deep syntactic roles such as ‘Actor’ (ACT), ‘Patient’ (PAT), ‘Addressee’ (ADDR), ‘Effect’ (EFF), ‘Direction’, ‘Location’, ‘Manner’, etc. Further, the information on obligatoriness (‘?’ in front of a role label indicates its optionality in this text) and on possible morphological forms (here in subscript) is specified for each valency complementation. Each lexical unit can be further described by additional syntactic and syntactic-semantic information, e.g., on reciprocity, diatheses (as e.g. passivization), syntactico-semantic class etc.

For the lexicographic representation of CPs, the VALLEX lexicon was extended to cover also predicative nouns. In addition, the respective lexical units representing light verbs and predicative nouns were enriched with attributes that allow a user to derive valency structures of the whole CPs – these attributes are thoroughly described in Section 4.

Grammar component. The grammar component represents a part of the overall grammar of Czech, it stores formal rules directly related to the valency structure of verbs. This component serves for an economic description of systematic changes in the valency structure of verbs associated with various syntactic phenomena, esp. with passivization and reciprocity. It also comprises rules allowing for the derivation of deep and surface syntactic structures of CPs. These rules are presented in Section 3.

3 Grammar Component: Formation of Deep and Surface Syntactic Structures of CPs

3.1 Deep Syntactic Structure

The deep syntactic structure of CPs is formed by both valency complementations from the valency frame of the light verb and complementations from the frame of the predicative noun.

Predicative nouns. The valency frame of a predicative noun describes the usage of the noun in nominal structures. Individual valency complementations are semantically saturated – they correspond to individual semantic participants characterizing a situation denoted by the noun, as can be exemplified on the predicative noun dohoda, ‘agreement’, see its valency frame and example illustrating its nominal structure in (3) and the correspondence between its valency complementations and its semantic participants in (4):

\[
\begin{align*}
(3) \quad \text{dohoda}_\text{ps} & \quad \text{‘agreement’} \\
\quad \text{ACT}_\text{2,ps} & \quad \text{ADDR}_\text{ps,7} & \quad \text{PAT}_\text{až} & \quad \text{inf}_\text{až} & \quad \text{obý,addr,ls,cont} \\
\quad \text{dohoda Francie}_\text{party,1} & \quad \text{ACT}_2 \\
\quad \text{s Německem}_\text{party,2} & \quad \text{ADDR}_\text{a7} & \quad \text{Obligation}_\text{pat,5} & \quad \text{obý,addr,ls,cont} \\
\quad \text{‘the agreement of France with Germany on non-aggression’} \\
\end{align*}
\]

\[
\text{ACT} \iff \text{Party}_1 \\
\text{ ADDR} \iff \text{Party}_2 \\
\text{PAT} \iff \text{Obligation}
\]

Light verbs. The deep structure of a light verb is formed by its valency frame, with one position (labeled CPHR) reserved for a predicative noun. A single light verb may be characterized by different deep syntactic structures, i.e., described by different valency frames which combine with different predicative nouns, see e.g. the light verb uzavřít, in (5) and (7).

1 http://lindat.mff.cuni.cz/services/PDT-Vallex/
2 http://ufal.mff.cuni.cz/vallex/3.0/
3 http://ufal.mff.cuni.cz/pdt3.0
Light verbs, being (to some extent) semantically bleached, do not evoke any semantic participants. As a result, their valency complementations are characterized primarily as semantically underspecified deep syntactic positions, see schemes provided in (6) and (8) (compare also with [5]).

(5) uzavřít<sub>LV</sub> ‘make’
   ACT, ADDR, CPHR,
   ACT ⇔ 0
(6) ADDR ⇔ 0
   CPHR ⇔ PN
   (This LV combines, e.g., with the PNs dohoda ‘agreement’ and sázka ‘bet.’)
(7) uzavřít<sub>LV</sub> ‘end, conclude’:
   ACT, CPHR,
   ACT ⇔ 0
   CPHR ⇔ PN
   (This LV combines, e.g., with the PNs debata ‘discussion’ and výstěřování ‘inquiry’.)

The only exception when a light verb contributes its semantic participant is represented by CPs with causative LVs. The causative LVs are seen as initiating the event denoted by the predicative noun selecting the given verb. These verbs thus contribute the ‘Instigator’ participant (and the nouns their respective semantic participants). For example, the LV uzavřít<sub>LV</sub> ‘close’ that is instantiated, e.g., in the CP uzavřít<sub>ACT</sub> přístup ‘close an access’ represents the causative LV, with the ‘Instigator’ mapped onto its ACT, see the valency frame of this verb (9) and the scheme of the mapping of semantic participants and valency complementations (10):

(9) uzavřít<sub>LV</sub> ‘close’:
   ACT, CPHR, ‘BEN’,
   ACT ⇔ Instigator
   CPHR ⇔ PN
   BEN ⇔ 0
   (This LV combines, e.g., with the PN přístup ‘access’.)

Within CPs, semantically underspecified valency complementations of LVs acquire semantic capacity via coreference with valency complementations of the predicative nouns with which they form CPs. These coreferential relations between valency complementations of LVs and complementations of PNs thus characterize the deep syntactic structure of individual CPs.

Complex predicates with light verbs. The deep syntactic structure of a CP is formed via an interplay between the valency frames of the respective LV and PN that form the given CP. A crucial role in the formation of the deep syntactic structure of a CP plays (i) the number of semantic participants involved in a situation denoted by the CP, and (ii) coreferential relations between the valency complementations of the LV and the PN [8]. The deep syntactic structure of a CP thus consists of:

- all nominal valency complementations, as they (directly) correspond to semantic participants;
- all verbal valency complementations, as their semantic saturation is acquired in one of the following ways:
  - the CPHR valency position, as it is reserved for the predicative noun;
  - the verbal valency complementation corresponding to the ‘Instigator’ participant (if present);
  - other verbal valency complementations, as they corefer with individual nominal valency complementations.

Let us exemplify the deep structure formation on the example of the CP uzavřít dohodu ‘make an agreement’. The predicative noun dohodžení ‘agreement’ is characterized by three semantic participants corresponding to three valency complementations of this noun, as indicated in (3) and (4). The light verb uzavřít<sub>LV</sub> ‘make’ is characterized by the valency frame provided in (5). The CPHR position of the light verb is filled with the PN dohoda ‘agreement’, the remaining valency complementations ACT and ADDR of the light verb enter into coreference with the ACT and ADDR of the given predicative noun, respectively (thus they obtain their semantic capacity from the given nominal complementations), see scheme (11), the sentence below and the deep dependency tree of the given CP in Figure 1:

(11) uzavřít dohodu ‘make an agreement’:
   ACT<sub>LV</sub> ⇔ ACT<sub>PN</sub> ⇔ Party<sub>1</sub>
   ADDR<sub>LV</sub> ⇔ ADDR<sub>PN</sub> ⇔ Party<sub>2</sub>
   PAT<sub>PN</sub> ⇔ Obligation
   CPHR<sub>LV</sub> ⇔ dohodžení

Francie<sub>PN</sub> uzavřela s Německem<sub>PN</sub> dohodžení o neútočení; Obligation.
‘France made an agreement with Germany on non-aggression.’

In many cases, a predicative noun can select different light verbs (and thus create different CPs), and so makes it possible to embed the expressed event “into different general semantic scenarios and thus to perspectivize it from the point of view of different participants” [8]. For example, the predicative noun rozkázání

---

4In the schemes, correspondence between semantic participants and valency complementations is marked with ⇔ whereas ↔ is reserved for coreference relations.
selects either the light verb dostat ‘to give’, or the light verb dostat dohodu ‘to get’. This noun evokes three semantic participants, namely ‘Speaker’, ‘Recipient’, and ‘Information’. When it selects the light verb dostat ‘to give’, the situation expressed by this noun is viewed from the perspective of the ‘Speaker’ as it occupies the prominent subject position given by the ACT of the light verb, see example (12), while selecting the light verb dostat dohodu ‘to get’, the situation is perspectivized from the ‘Recipient’, see example (13).

(12) Generálu pakedor.AKT.-LV dal rozkaz
vojákmü_bopos.Inform.ATT.-LV k ústupu

‘The general gave soldiers the order to retreat.’

(13) Vojácumü_bopos.ORG.-LV do generálu
rozkaz k ústupu

‘Soldiers got the order to retreat by the general.’

3.2 Surface Syntactic Structure

The theoretical analysis supported by the extensive empirical data annotation has revealed that with CPs in Czech, each semantic participant is typically expressed in the surface sentence just once. Despite the fact that semantic participants are contributed to CPs – with the exception of the verbal ‘Instigator’ – by predicative nouns, Czech CPs have a strong tendency to express these participants in the surface structure as verbal modifications, see as well [7]. Namely, those participants characterizing a CP that are referred to by both valency complementations of the PN as well as (via coreference) complementations of the LV are primarily expressed on the surface as verbal modifications. On the other hand, those participants that are mapped onto valency complementations of the PN are realized as the nominal modifications.

As a result, the rules governing the formation of the surface syntactic structure of Czech CPs can be summarized as follows:

- As verbal modifications, all valency complementations from the valency frame of the light verb are primarily expressed in the surface structure, namely:
  (i) the valency complementation filled by the predicative noun (the CPHR functor): it is obligatorily expressed in the surface structure as a verbal modification;
  (ii) the valency complementation corresponding to ‘Instigator’ (if present): it can be expressed in the surface structure only as a verbal modification;
  (iii) other verbal valency complementations: they are primarily expressed in the surface structure as verbal modifications, too.
- As nominal modifications:
  (iv) those valency complementations from the valency frame of the predicative noun that are not in coreference with verbal ones are primarily expressed in the surface structure.6

For instance, within the CP uzavřít dohodu ‘make an agreement’, the following valency complementations are expressed in the surface structure: all the valency complementations of the LV uzavřít ‘to make’ (see its valency frame in (5)) are expressed as verbal modifications on the surface, namely: CPHR reserved for the predicative noun dohoda ‘agreement’ (principle (i) in the direct object position, the verbal ACT and ADDR in the subject position and the indirect object position, respectively (principle (iii)) these valency complementations refer to the ‘Party_1’ and ‘Party_2’ via coreference with the ACT and ADDR of the PN, see scheme (11)). From the valency complementations of the PN dohoda ‘agreement’ (the valency frame in (3)), only PAT (referring to ‘Obligation’, not being in coreference with any verbal complementation) is expressed on the surface as a nominal modification (principle (iv)); the remaining ACT and ADDR complementations of this noun (being in coreference with the verbal ACT and ADDR) are subject to systemic ellipsis; see the example sentence below and its surface dependency tree in Figure 2:

Francíkparty1.AKT.-LV uzavřela s Německemparty2.AKT.-LV dohodou, CPHR.-LV o neútočení, Obligation.PAT.-PN.

---

6We disregard the cases of valency complementations unexpressed on the surface due to their optionality, actual ellipsis, generalization etc.

7In some cases, a nominal valency complementation corefering with a verbal one may be alternatively expressed in the surface structure as a nominal modification, see e.g.
(a) S Německemparty1.AKT.-LV Francíkparty2.AKT.-LV uzavřela dohodou, CPHR.-LV.
(b) Francíkparty1.AKT.-LV uzavřela dohodou, CPHR.-LV s Německemparty2.AKT.-PN.

---

5The only exception is represented by the semantic participant mapped onto the nominal ACT: under certain conditions, this participant can be expressed twice, both as a verbal and as a nominal modification (e.g., Vrchní komisařparty1.AKT.-LV žádá generálu, CPHR.-PAR.-PN. – vjezdovod, CPHR.-PART_2.PAR.-PN. uzavřel, pn. ‘The chief inspector has already concluded his investigation of the crime.’).
As was shown above, the deep and surface syntactic structures of CPs are formed as a combination of valency structures of respective predicative nouns and light verbs, with respect to the coreference between their individual valency complementations. The process of both the deep and surface structure formation is regular enough to be described on the rule basis. These rules operate on the information provided by the data component of the lexicon.

In the data component of the VALLEX lexicon, individual lexical units of verbs and predicative nouns are described. In addition to the core valency information in a form of valency frames, these lexical units carry three special attributes linking the respective pairs of lexical units of the PN and LV allowing for the derivation of both deep and surface syntactic structures of the whole complex predicate, namely attributes lvc, map and instig.

Attributes lvc. Respective lexical units of LVs and PNs that form CPs are linked by the attribute lvc, the value of which is a list of references to respective lexical units. This attribute is attached to lexical units of predicative nouns and (for user’s convenience) to lexical units of light verbs as well. Figure 3 illustrates three lexical units for the LV uzavřít, which are instantiated, e.g., in the CPs uzavřít dohodu ‘make an agreement’, uzavřít vyšetřování ‘close an investigation’, and uzavřít přístup ‘close an access’, respectively (simplified).

Attributes map. The information on the coreference between valency complementations of LVs and complementations of PNs is provided in the attribute map. This attribute is attached to PNs (as it is the PN that selects an appropriate LV). The value of the map attribute is a list of pairs of coreferring complementations. Figure 4 illustrates three lexical units for three PNs, namely, dohoda ‘agreement’ (see also (3)), vyšetřování ‘investigation’, and přístup ‘access’.

Each PN can be assigned more than one attribute map reflecting different coreference relations; in such cases, the map attributes are co-indexed with the relevant lvc attributes to allow for the correct formation of the CPs structures.

Attribute instig. The information on the mapping of the ‘Instigator’ onto a valency complementation of relevant LVs is recorded in the attribute instig attached to the verbal valency frame, see lexical unit 3 in Figure 3.

If a LV forms syntactic structures with different PNs characterized by different coreferential relations, the instig attribute is co-indexed with the respective lvc attribute, containing the list of references to PNs that select the LV with the ‘Instigator’.

The following Tables 1 and 2 summarize the corpus analysis of Czech CPs formed by 129 verb lemmas from the VALLEX lexicon (those LVs were selected that have at least one valency frame with the CPHR functor in the PDT corpus, see Section 2). The CPs were extracted from the Czech National Corpus.
SYN2010, by the Word Sketch Engine [17] allowing to identify for each verb lemma its nominal collocates expressed as its direct object (function has_obj4). From the obtained list of collocations, only those nominal collocates were indicated by human annotators that represent PNs (560 noun lemmas in total). As a key criterion for identifying CPs, the coreference between the ACT of the noun and some of valency complementations of the LV has been adopted [18]. This criterion was satisfied by 1,025 collocations, which represent the most frequent and semantically salient CPs of the selected light verbs.

The identified CPs were further annotated with respect to the coreference between valency complementations of the LV and PN and with respect to the mapping of ‘Instigator’ (where it was relevant), see esp. [10]. Tables 1 and 2 summarize results of the annotation process. Table 1 contains those CPs the light verbs of which behave unambiguously with respect to the causative feature, i.e., they are either non-causative (0 in the ‘Instig’ column), or causative. With the CPs with causative light verbs, the Instigator was mapped either onto verbal ACT, or onto verbal ORIG. In the annotation, 12 types of coreference relations between verbal and nominal valency complementations were identified: the most frequent was represented by the coreference between ACT of the light verb and ACT of the predicative noun (506 CPs, i.e. almost 50% of all analyzed CPs).

In the annotation, a specific type of CPs characterized by an ambiguous character with respect to the causativity of LVs was found (122 cases, i.e. almost 12% of CPs). These CPs are formed by PNs characterized by the semantic participants ‘Experimenter’ and ‘Stimulus’. Two situations occur with these CPs. First, the valency complementation of a PN corresponding to ‘Stimulus’ enters in coreference with ACT of the LV with which the given PN forms the CP (as exemplified in (14), Figure 5); in this case, the LV behaves as non-causative verb. Second, the given complementation of a PN is not in coreference with any verbal complementation; in this case, the LV contributes the ‘Instigator’ to the CP (example (15), Figure 6). For example, with the CP vyvolat protest, the semantic participant ‘Stimulus’ given by the PN protest_pn ‘protest’ mapped onto PAT of the noun either enters in coreference with the ACT of the LV vyvolat_lv ‘to raise’, see example (14), or remains without coreference, see example (15).

14 Stavba[stimulator]ACT-LV dálnice vyvolala u obyvatel[experimenter]LOC-LV protesty[PN,CPHR-LV]. 'The construction of the motorway has prompted protests of the inhabitants.'

15 Stavba[instigator]ACT-LV dálnice vyvolala u obyvatel[experimenter]LOC-LV protesty[PN,CPHR-LV proti postupu[stimulator,PAT-PN] radních. 'The construction of the motorway has prompted protests of the inhabitants against the decision of councillors.'
Table 1: Unambiguous Czech CPs identified in the corpus data, sorted according to causativity of LVs and types of coreference between verbal and nominal valency complementations.

<table>
<thead>
<tr>
<th>‘Instig’ coreference</th>
<th>#</th>
<th>%</th>
<th>examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0) ACT&lt;sub&gt;pn&lt;/sub&gt;− ACT&lt;sub&gt;lv&lt;/sub&gt;</td>
<td>506</td>
<td>49.4</td>
<td>mít chuť, vést život, uzavřít debatu, uzavřít vyšetřování</td>
</tr>
<tr>
<td>ACT&lt;sub&gt;pn&lt;/sub&gt;− ACT&lt;sub&gt;lv&lt;/sub&gt; &amp; ADDR&lt;sub&gt;pn&lt;/sub&gt;− ADDR&lt;sub&gt;lv&lt;/sub&gt;</td>
<td>120</td>
<td>11.7</td>
<td>dát rozkaz, poskytnout rozhovor, uzavřít dohodu, uzavřít sázku</td>
</tr>
<tr>
<td>ACT&lt;sub&gt;pn&lt;/sub&gt;− ACT&lt;sub&gt;lv&lt;/sub&gt; &amp; PAT&lt;sub&gt;pn&lt;/sub&gt;− ADDR&lt;sub&gt;lv&lt;/sub&gt;</td>
<td>93</td>
<td>9.1</td>
<td>navázat vztah</td>
</tr>
<tr>
<td>ACT&lt;sub&gt;pn&lt;/sub&gt;− ORIG&lt;sub&gt;lv&lt;/sub&gt; &amp; ADDR&lt;sub&gt;pn&lt;/sub&gt;− ACT&lt;sub&gt;lv&lt;/sub&gt;</td>
<td>28</td>
<td>2.7</td>
<td>dostat nabídku, získat informace</td>
</tr>
<tr>
<td>ACT&lt;sub&gt;pn&lt;/sub&gt;− ORIG&lt;sub&gt;lv&lt;/sub&gt; &amp; PAT&lt;sub&gt;pn&lt;/sub&gt;− ACT&lt;sub&gt;lv&lt;/sub&gt;</td>
<td>22</td>
<td>2.1</td>
<td>dostat rán, dostat pokutu</td>
</tr>
<tr>
<td>ACT&lt;sub&gt;pn&lt;/sub&gt;− ACT&lt;sub&gt;lv&lt;/sub&gt; &amp; PAT&lt;sub&gt;pn&lt;/sub&gt;− DIR3&lt;sub&gt;lv&lt;/sub&gt;</td>
<td>28</td>
<td>2.7</td>
<td>obraceť pozornost, položit důraz</td>
</tr>
<tr>
<td>ACT&lt;sub&gt;pn&lt;/sub&gt;− ACT&lt;sub&gt;lv&lt;/sub&gt; &amp; PAT&lt;sub&gt;pn&lt;/sub&gt;− LOC&lt;sub&gt;lv&lt;/sub&gt;</td>
<td>22</td>
<td>2.1</td>
<td>najít inspiraci, najít potěšení</td>
</tr>
<tr>
<td>ACT&lt;sub&gt;pn&lt;/sub&gt;− LOC&lt;sub&gt;lv&lt;/sub&gt; &amp; PAT&lt;sub&gt;pn&lt;/sub&gt;− ACT&lt;sub&gt;lv&lt;/sub&gt;</td>
<td>22</td>
<td>2.1</td>
<td>najít odevz, nalézt pochopení</td>
</tr>
<tr>
<td>ACT&lt;sub&gt;lv&lt;/sub&gt; ACT&lt;sub&gt;pn&lt;/sub&gt;− ADDR&lt;sub&gt;lv&lt;/sub&gt;</td>
<td>53</td>
<td>5.2</td>
<td>dát nadvě, vynést slávu, vzít odevahu</td>
</tr>
<tr>
<td>ACT&lt;sub&gt;pn&lt;/sub&gt;− LOC&lt;sub&gt;lv&lt;/sub&gt;</td>
<td>26</td>
<td>2.5</td>
<td>prohouzet podezřívanost, vzùdít zdání</td>
</tr>
<tr>
<td>ACT&lt;sub&gt;pn&lt;/sub&gt;− BEN&lt;sub&gt;lv&lt;/sub&gt;</td>
<td>8</td>
<td>0.8</td>
<td>zvednout náladu, otevír přístop, uzavřít přístop</td>
</tr>
</tbody>
</table>

Table 2: Ambiguous Czech CPs identified in the corpus data, sorted according to causativity of LVs and types of coreference between verbal and nominal valency complementations.

<table>
<thead>
<tr>
<th>without ‘Instigator’ coreference</th>
<th>with ‘Instigator’ coreference</th>
<th>#</th>
<th>%</th>
<th>examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACT&lt;sub&gt;pn&lt;/sub&gt;− LOC&lt;sub&gt;lv&lt;/sub&gt; &amp; PAT&lt;sub&gt;pn&lt;/sub&gt;− ACT&lt;sub&gt;lv&lt;/sub&gt;</td>
<td>ACT&lt;sub&gt;lv&lt;/sub&gt; ACT&lt;sub&gt;pn&lt;/sub&gt;− LOC&lt;sub&gt;lv&lt;/sub&gt;</td>
<td>92</td>
<td>9.0</td>
<td>vyvolat protest, budít důvěru</td>
</tr>
<tr>
<td>ACT&lt;sub&gt;pn&lt;/sub&gt;− ADDR&lt;sub&gt;lv&lt;/sub&gt; &amp; PAT&lt;sub&gt;pn&lt;/sub&gt;− ACT&lt;sub&gt;lv&lt;/sub&gt;</td>
<td>ACT&lt;sub&gt;lv&lt;/sub&gt; ACT&lt;sub&gt;pn&lt;/sub&gt;− ADDR&lt;sub&gt;lv&lt;/sub&gt;</td>
<td>23</td>
<td>2.2</td>
<td>přinést radost, činit obtíž</td>
</tr>
<tr>
<td>ACT&lt;sub&gt;pn&lt;/sub&gt;− LOC&lt;sub&gt;lv&lt;/sub&gt; &amp; ORIG&lt;sub&gt;pn&lt;/sub&gt;− ACT&lt;sub&gt;lv&lt;/sub&gt;</td>
<td>ACT&lt;sub&gt;lv&lt;/sub&gt; ACT&lt;sub&gt;pn&lt;/sub&gt;− LOC&lt;sub&gt;lv&lt;/sub&gt;</td>
<td>3</td>
<td>0.3</td>
<td>vzbdít pocit</td>
</tr>
<tr>
<td>ACT&lt;sub&gt;pn&lt;/sub&gt;− LOC&lt;sub&gt;lv&lt;/sub&gt; &amp; ADDR&lt;sub&gt;pn&lt;/sub&gt;− ACT&lt;sub&gt;lv&lt;/sub&gt;</td>
<td>ACT&lt;sub&gt;lv&lt;/sub&gt; ACT&lt;sub&gt;pn&lt;/sub&gt;− LOC&lt;sub&gt;lv&lt;/sub&gt;</td>
<td>4</td>
<td>0.4</td>
<td>vyvolat podezření</td>
</tr>
</tbody>
</table>

6 Conclusion

In this paper, we have summarized results of our analysis of Czech complex predicates with light verbs. We have described its lexicographic model based on a close cooperation of the lexical and grammar component. Although our proposal is primarily designed for the Valency Lexicon of Czech verbs VALLEX, we suppose that its main tenets can be easily adopted by other lexical resources as well. Finally, we have introduced the annotation of a large collection of linguistic data which will be integrated in the VALLEX lexicon soon.

Acknowledgements

The work on this project has been supported by the grant of the Czech Science Foundation (project GA15-09979S) and partially by the LINDAT/CLARIN project of the Ministry of Education, Youth and Sports of the Czech Republic (project LM2015071).

This work has been using language resources distributed by the LINDAT/CLARIN project of the Ministry of Education, Youth and Sports of the Czech Republic (project LM2015071).

References


