

# Complex Predicates with Light Verbs in VALLEX: From Formal Model to Lexicographic Description

Václava Kettnerová and Markéta Lopatková

Charles University, Faculty of Mathematics and Physics, Prague, Czech Republic  
{kettnerova, lopatkova}@ufal.mff.cuni.cz

*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* ‘to have a fun’, or *uzavřít dohodu* ‘make an agreement’. Due to their complex characteristics, CPs proven to be challenging for syntactic theories as well as to 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 participants are provided by the predicative noun *dohoda* ‘agreement’. However, two of these participants – ‘Party\_1’ and ‘Party\_2’ – are expressed in the surface structure of the sentence not as nominal but as verbal modifications, namely as the subject and as the indirect object, while only the participant ‘Obligation’ is expressed as a nominal modification, namely as its attribute, see (1). The syntactic structure of the given

sentence is thus formed by valency complementations of both the light verb and the predicative noun. In contrast, the sentence with the predicative verb *uzavřít* ‘close; turn off’, see e.g. (2), is characterized by two participants, ‘Agent’ and ‘Affected\_object’, being evoked by the verb, they are expressed on the surface as valency complementations of the given verb.

- (1) *Francie*<sub>Sb-verb</sub> *uzavřela s Německem*<sub>InObj-verb</sub> *dohodu*<sub>Obj-verb</sub> *o neútočení*<sub>Atr-noun</sub>.  
‘France made an agreement with Germany on non-aggression.’
- (2) *Hasiči*<sub>Sb-verb</sub> *uzavřeli přívod*<sub>Obj-verb</sub> *plynu*.  
‘Firemen turned off the gas main.’

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, 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 theory of FGD has been applied in several valency lexicons, esp. PDT-Vallex<sup>1</sup> [13] and VALLEX<sup>2</sup> [14], and verified on extensive corpus data, esp. within the Prague Dependency Treebank (PDT)<sup>3</sup>. 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 allows 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 or reciprocity. It also comprises rules allowing for the derivation of deep and surface syntactic structure of CPs. These rules are presented in Section 3.

## 3 Grammar Component: Formation of Deep and Surface Syntactic Structure

### 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*<sub>PN</sub> ‘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):

- (3) *dohoda*<sub>PN</sub> ‘agreement’:  
 ACT<sub>2,pos</sub> ADDR<sub>s+7</sub> PAT<sub>na+6,o+6,inf,aby,zda,e,cont</sub>  
*dohoda* Francie<sub>Party\_1,ACT(2)</sub>  
*s Německem*<sub>Party\_2,ADDR(s+7)</sub> *o neútočení*<sub>Obligation,PAT(o+6)</sub>  
 ‘the agreement of France with Germany  
 on non-aggression’
- (4) ACT ⇔ Party\_1  
 ADDR ⇔ Party\_2  
 PAT ⇔ 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, depending on a predicative noun with which it forms CPs, see e.g. the light verb *uzavřít*<sub>LV</sub> in (5) and (7).

<sup>1</sup><http://lindat.mff.cuni.cz/services/PDT-Vallex/>

<sup>2</sup><http://ufal.mff.cuni.cz/vallex/3.0/>

<sup>3</sup><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 typically semantically unsaturated; in other words, the 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<sub>1</sub> ADDR<sub>s+7</sub> CPHR<sub>4</sub>  
 ACT ⇔ ∅
- (6) ADDR ⇔ ∅  
 CPHR ⇔ PN  
 (This LV combines, e.g., with the PNs *dohoda* ‘agreement’, *sázka* ‘bet’.)
- (7) *uzavřít<sub>LV</sub>* ‘end, conclude’:  
 ACT<sub>1</sub> CPHR<sub>4</sub>
- (8) ACT ⇔ ∅  
 CPHR ⇔ PN  
 (This LV combines, e.g., with the PNs *debata* ‘discussion’, *vyšetřování* ‘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 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<sub>1</sub> CPHR<sub>4</sub> ?BEN<sub>3</sub>  
 ACT ⇔ Instigator
- (10) CPHR ⇔ PN  
 BEN ⇔ ∅  
 (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 *dohoda<sub>PN</sub>* ‘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 PAT of the given predicative noun, respectively (thus they obtain their semantic capacity from the given nominal complementations), see scheme (11) and the sentence below with its deep dependency tree in Figure 1:

- (11) *uzavřít dohodu* ‘make an agreement’:<sup>4</sup>  
 ACT<sub>LV</sub> ↔ ACT<sub>PN</sub> ⇔ Party<sub>1</sub>  
 CPHR<sub>LV</sub> ⇔ dohoda<sub>PN</sub>  
 ADDR<sub>LV</sub> ↔ ADDR<sub>PN</sub> ⇔ Party<sub>2</sub>  
 PAT<sub>PN</sub> ⇔ Obligation

*Francie<sub>Party<sub>1</sub></sub> uzavřela s Německem<sub>Party<sub>2</sub></sub> dohodu<sub>PN</sub> o neútočení<sub>Obligation</sub>.*

‘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 *rozkaz<sub>PN</sub>*

<sup>4</sup>In the schemes, correspondence between semantic participants and valency complementations is marked with ⇔ whereas ↔ is reserved for coreference relations.

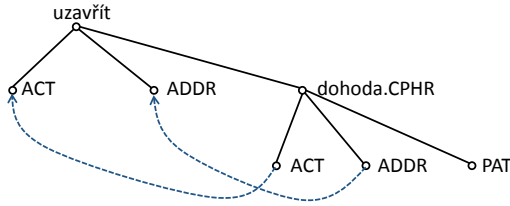


Figure 1: The deep dependency structure of the CP *uzavřít dohodu* ‘make an agreement’.

selects either the light verb *dát<sub>LV</sub>* ‘to give’, or the light verb *dostat<sub>LV</sub>* ‘to get’. This noun evokes three semantic participants, namely ‘Speaker’, ‘Recipient’, and ‘Information’. When it selects the light verb *dát<sub>LV</sub>* ‘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<sub>LV</sub>* ‘to get’, the situation is perspectivized from the ‘Recipient’, see example (13).

(12) *Generál<sub>Speaker,ACT-LV</sub> dal rozkaz vojákům<sub>Recipient,ADDR-LV</sub> k ústupu<sub>Information,PAT-PN</sub>.*  
 ‘The general gave soldiers the order to retreat.’

(13) *Vojáci<sub>Recipient,ACT-LV</sub> dostali od generála<sub>Speaker,ORIG-LV</sub> rozkaz k ústupu<sub>Information,PAT-PN</sub>.*  
 ‘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.<sup>5</sup> 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 the verbal modifications. On the other hand, those participants that are mapped only onto valency complementations of the PN are realized as the nominal modifications. As

<sup>5</sup>The 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 komisař<sub>Agens,ACT(1)-LV</sub> již své<sub>Agens,ACT(pos)-PN</sub> vyšetřování<sub>PN</sub> zločinu<sub>Incident,PAT(2)-PN</sub> uzavřel<sub>LV</sub>.* ‘The chief inspector has already concluded his investigation of the crime.’).

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

- All valency complementations from the *valency frame of the light verb* are primarily expressed in the surface structure, namely:<sup>6</sup>
  - (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, those valency complementations from the *valency frame of the predicative noun* are primarily expressed in the surface structure that are not in coreference with verbal ones.<sup>7</sup>

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 modification on the surface, namely, CPHR reserved for the predicative noun *dohoda* ‘agreement’ (principle (i)), the verbal ACT and ADDR (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:

*Francie<sub>Party,1,ACT-LV</sub> uzavřela s Německem<sub>Party,2,ADDR-LV</sub> dohodu<sub>PN,CPHR-LV</sub> o neútočení<sub>Obligation,PAT-PN</sub>.*

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

<sup>7</sup>In 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 Německem<sub>Party,2,ACT-LV</sub> Francie<sub>Party,1,ACT-LV</sub> uzavřela dohodu<sub>PN,CPHR-LV</sub>.* vs. (b) *Francie<sub>Party,1,ACT-LV</sub> uzavřela dohodu<sub>PN,CPHR-LV</sub> s Německem<sub>Party,2,ACT-CN</sub>,* with the ‘Party\_2’ participant (*s Německem*) preferably analyzed as a verbal (in (a)) or a nominal (in (b)) modification.

‘France made an agreement with Germany on non-aggression.’

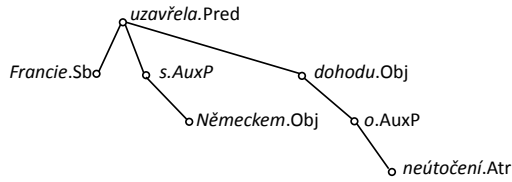


Figure 2: The surface dependency structure of the sentence *Francie uzavřela s Německem dohodu o neútočení*. ‘France made an agreement with Germany on non-aggression.’ (simplified)

## 4 Data Component: Interlinking Lexical Units

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 the deep and surface syntactic structures of the whole complex predicate, namely attributes *lvc*, *map* and *instig*.

**Attribute *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<sub>LV</sub>* ‘make; end, conclude; close, terminate’ (see also (5), (7) and (9)).

If a LV forms different syntactic structures with different PNs, more instances of the attribute *lvc* (indexed with numbers) are assigned to the relevant lexical unit.

**Attribute *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 corefering complementations. Figure 4 illustrates three lexical units for three PNs *dohoda<sub>PN</sub>* ‘agreement’ (see also (3)), *vyšetřování<sub>PN</sub>* ‘investigation’, and *přístup<sub>PN</sub>* ‘access’.

Each PN can be assigned with 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 the LV is recorded in the attribute *instig* attached to the verbal valency frame, see Figure 3.

If a LV forms different syntactic structures with different PNs, 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’.

| uzavírat <sup>impf</sup> , uzavřít <sup>pf</sup> |  |
|--|--|
| ①  | LV (impf: sjednávat; pf: sjednat) ‘make’<br>-frame: ACT <sub>1</sub> ADDR <sub>3,7</sub> CPHR <sub>4</sub><br>-example: <i>Firmy uzavíraly s Lucemburskem tajné dohody.</i><br>-lvc: dohoda-1, kompromis-1, ‘agreement, compromise,<br>kontrakt-1, obchod-1, contract, trade,<br>pakt-1, sázka-1, smlouva-1 pact, bet, contract’ |
| ②  | LV (impf: ukončovat; pf: ukončit) ‘end; conclude’<br>-frame: ACT <sub>1</sub> CPHR <sub>4</sub><br>-example: <i>Policie uzavírá vyšetřování všech tří případů.</i><br>-lvc: debata-1, vyšetřování-1 ‘discussion, investigation’  |
| ③  | LV (impf: zamezovat; pf: zamezit) ‘close; end, terminate’<br>-frame: ACT <sub>1</sub> CPHR <sub>4</sub> ?BEN <sub>3</sub><br>-example: <i>Dohoda ale uzavírá přístup na hranici.</i><br>-lvc: přístup-1 ‘access’<br>-instig: ACT   |
| ...  |  |

Figure 3: 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).

## 5 Corpus Data Analysis

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

|   |
|---|
| <p><b>dohoda</b></p> <p>① ujednání; domluva 'agreement'</p> <p>-frame: ACT<sub>2,pos</sub> ADDR<sub>s+7</sub> PAT<sub>na+6,o+6,inf,aby,zda,že,cont</sub></p> <p>-example: <i>dohoda Francie s Německem o neútočení</i></p> <p>-lvc: uzavírat/uzavřít-1, vypovídat/vypovědět-5</p> <p>-map: ACT<sub>PN</sub>-ACT<sub>LV</sub> &amp; ADDR<sub>PN</sub>-ADDR<sub>LV</sub></p> <p>...</p> |
| <p><b>vyšetřování</b></p> <p>① objasňování; prozkoumávání 'investigation'</p> <p>-frame: ACT<sub>2,pos</sub> PAT<sub>2,pos</sub></p> <p>-example: <i>vyšetřování všech odhalených případů zpronevěry</i></p> <p>-lvc: uzavírat/uzavřít-2, vést-5</p> <p>-map: ACT<sub>PN</sub>-ACT<sub>LV</sub></p> <p>...</p>  |
| <p><b>přístup</b></p> <p>① možnost někam vstoupit; přistoupení 'access'</p> <p>-frame: ACT<sub>1</sub> DIR<sub>3</sub><sub>do+2,k+3,na+4</sub></p> <p>-example: <i>přístup na hranici; přístup na trh práce</i></p> <p>-lvc: otvírat/otevírat/otevřít-1, uzavírat/uzavřít-3</p> <p>-map: ACT<sub>PN</sub>-BEN<sub>LV</sub></p> <p>...</p>   |

Figure 4: Three lexical units for the PNs *dohoda* ‘agreement’, *vyšetřování* ‘investigation’, and *přístup* ‘access’, respectively (simplified).

to identify for each verb lemma its nominal collocates expressed as 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 ( $\emptyset$  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 coreferential relation 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 was found (122 cases, i.e. almost 12%

of CPs). These CPs are formed by PNs characterized by the semantic participants ‘Experiencer’ and ‘Stimul’. Two situations occur with these CPs. First, the valency complementation of a PN corresponding to ‘Stimul’ 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 behave as a 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 ‘Stimul’ given by the PN *protest*<sub>PN</sub> ‘protest’ mapped onto PAT of the noun either enters in coreference with the ACT of the LV *vyvolat*<sub>LV</sub> ‘to raise’, see example (14), or remains without coreference, see example (15).

(14) *Stavba*<sub>Stimul,ACT-LV</sub> *dálnice* *vyvolala*  
*u obyvatel*<sub>Experiencer,LOC-LV</sub> *protesty*<sub>PN,CPHR-LV</sub>.  
‘The construction of the motorway has prompted protests of the inhabitants.’

(15) *Stavba*<sub>Instigator,ACT-LV</sub> *dálnice* *vyvolala*  
*u obyvatel*<sub>Experiencer,LOC-LV</sub> *protesty*<sub>PN,CPHR-LV</sub> *proti*  
*postupu*<sub>Stimul,PAT-PN</sub> *radních*.  
‘The construction of the motorway has prompted protests of the inhabitants against the decision of councillors.’

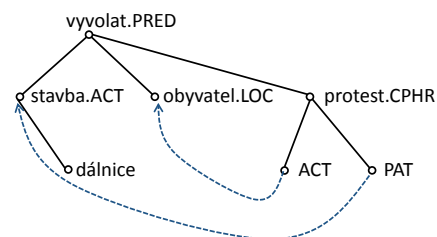


Figure 5: The deep dependency structure of the non-causative example (14) (simplified).

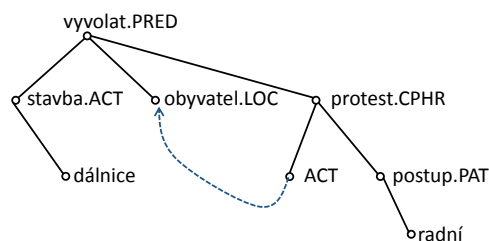


Figure 6: The deep dependency structure of the causative example (15) (simplified).

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.

| ‘Instig’           | coreference                           |  | #   | %    | examples   |
|--------------------|---------------------------------------|--|-----|------|--|
| $\emptyset$        | ACT <sub>PN</sub> –ACT <sub>LV</sub>  |  | 506 | 49.4 | <i>mít chut’, vést život, uzavřít debatu, uzavřít vyšetřování</i>    |
|                    | ACT <sub>PN</sub> –ACT <sub>LV</sub>  | & ADDR <sub>PN</sub> –ADDR <sub>LV</sub> | 120 | 11.7 | <i>dát rozkaz, poskytnout rozhovor uzavřít dohodu, uzavřít sázku</i> |
|                    | ACT <sub>PN</sub> –ACT <sub>LV</sub>  | & PAT <sub>PN</sub> –ADDR <sub>LV</sub>  | 93  | 9.1  | <i>navázat vztah</i>   |
|                    | ACT <sub>PN</sub> –ORIG <sub>LV</sub> | & ADDR <sub>PN</sub> –ACT <sub>LV</sub>  | 28  | 2.7  | <i>dostat nabídku, získat informace</i>                              |
|                    | ACT <sub>PN</sub> –ORIG <sub>LV</sub> | & PAT <sub>PN</sub> –ACT <sub>LV</sub>   | 22  | 2.1  | <i>dostat ránu, dostat pokutu</i>                                    |
|                    | ACT <sub>PN</sub> –ACT <sub>LV</sub>  | & PAT <sub>PN</sub> –DIR3 <sub>LV</sub>  | 28  | 2.7  | <i>obracet pozornost, položit důraz</i>                              |
|                    | ACT <sub>PN</sub> –ACT <sub>LV</sub>  | & PAT <sub>PN</sub> –LOC <sub>LV</sub>   | 22  | 2.1  | <i>najít inspiraci, najít potěšení</i>                               |
|                    | ACT <sub>PN</sub> –LOC <sub>LV</sub>  | & PAT <sub>PN</sub> –ACT <sub>LV</sub>   | 22  | 2.1  | <i>najít odezvu, nalézt pochopení</i>                                |
| ACT <sub>LV</sub>  | ACT <sub>PN</sub> –ADDR <sub>LV</sub> |  | 53  | 5.2  | <i>dát naději, vynést slávu, vzít odvahu</i>                         |
|                    | ACT <sub>PN</sub> –LOC <sub>LV</sub>  |  | 26  | 2.5  | <i>probouzet podezřavost, vzbudit zdání</i>                          |
|                    | ACT <sub>PN</sub> –BEN <sub>LV</sub>  |  | 8   | 0.8  | <i>zvednout náladu, otevřít přístup, uzavřít přístup</i>             |
| ORIG <sub>LV</sub> | ACT <sub>PN</sub> –ACT <sub>LV</sub>  |  | 18  | 1.8  | <i>dostat příležitost, získat výhodu</i>                             |

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.

| without ‘Instigator’<br>coreference   |   |                                       | with ‘Instigator’<br>‘Instig’ coreference |                                       | #  | %   | examples                             |
|---------------------------------------|---|---------------------------------------|---|---------------------------------------|----|-----|--------------------------------------|
| ACT <sub>PN</sub> –LOC <sub>LV</sub>  | & | PAT <sub>PN</sub> –ACT <sub>LV</sub>  | ACT <sub>LV</sub>                         | ACT <sub>PN</sub> –LOC <sub>LV</sub>  | 92 | 9.0 | <i>vyvolat protest, budit důvěru</i> |
| ACT <sub>PN</sub> –ADDR <sub>LV</sub> | & | PAT <sub>PN</sub> –ACT <sub>LV</sub>  | ACT <sub>LV</sub>                         | ACT <sub>PN</sub> –ADDR <sub>LV</sub> | 23 | 2.2 | <i>přinést radost, činit obtíž</i>   |
| ACT <sub>PN</sub> –LOC <sub>LV</sub>  | & | ORIG <sub>PN</sub> –ACT <sub>LV</sub> | ACT <sub>LV</sub>                         | ACT <sub>PN</sub> –LOC <sub>LV</sub>  | 3  | 0.3 | <i>vzbudit pocit</i>                 |
| ACT <sub>PN</sub> –LOC <sub>LV</sub>  | & | ADDR <sub>PN</sub> –ACT <sub>LV</sub> | ACT <sub>LV</sub>                         | ACT <sub>PN</sub> –LOC <sub>LV</sub>  | 4  | 0.4 | <i>vyvolat podezření</i>             |

## 6 Conclusion

In this paper, we have summarized results of our analysis of Czech complex predicates with light verbs. We describe 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

- [1] Algeo, J.: Having a look at the expanded predicate. In Aarts, B., Meyer, C.F., eds.: *The Verb in Contemporary English: Theory and Description*. Cambridge University Press, Cambridge (1995) 203–217
- [2] Grimshaw, J., Mester, A.: Light verbs and  $\theta$ -marking. *Linguistic inquiry* **19** (1988) 205–232
- [3] Butt, M.: The light verb jungle: Still hacking away. In Amberber, M., Baker, B., Harvey, M., eds.: *Complex Predicates in Cross-Linguistic Perspective*. Cambridge University Press, Cambridge (2010) 48–78

- [4] Hinrichs, E., Kathol, A., Nakazawa, T.: *Complex Predicates in Nonderivational Syntax*. Syntax and Semantics 30. Academic Press, San Diego (1998)
- [5] Alonso Ramos, M.: Towards the synthesis of support verb constructions: Distribution of syntactic actants between the verb and the noun. In Wanner, L., Mel'čuk, I.A., eds.: *Selected Lexical and Grammatical Issues in the Meaning-Text Theory*. John Benjamins Publishing Company, Amsterdam, Philadelphia (2007) 97–137
- [6] Radimský, J.: *Verbo-nominální predikát s kategoriálním slovesem*. Editio Universitatis Bohemiae Meridionalis, České Budějovice (2010)
- [7] Macháčková, E.: Constructions with verbs and abstract nouns in Czech (analytical predicates). In Čmejrková, S., Štícha, F., eds.: *The Syntax of Sentence and Text: A Festschrift for František Daneš*. Volume 42 of *Linguistic and Literary Studies in Eastern Europe*. John Benjamins Publishing Company, Amsterdam, Philadelphia (1994) 365–374
- [8] Kettnerová, V., Lopatková, M.: At the lexicon-grammar interface: The case of complex predicates in the functional generative description. In Hajičová, E., Nivre, J., eds.: *Proceedings of Depling 2015*, Uppsala, Sweden, Uppsala University (2015) 191–200
- [9] Kettnerová, V.: *Syntaktická struktura komplexních predikátů*. *Slovo a slovesnost* (2017) (in print).
- [10] Kettnerová, V., Lopatková, M.: *Ke koreferenci u komplexních predikátů s kategoriálním slovesem*. *Korpus – gramatika – axiologie* (2017) (submitted).
- [11] Sgall, P., Hajičová, E., Panevová, J.: *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects*. Reidel, Dordrecht (1986)
- [12] Panevová, J.: *Valency Frames and the Meaning of the Sentence*. In Luelsdorff, P.A., ed.: *The Prague School of Structural and Functional Linguistics*. John Benjamins Publishing Company, Amsterdam/Philadelphia (1994) 223–243
- [13] Uřešová, Z.: *Valence sloves v Pražském závislostním korpusu. Ústav formální a aplikované lingvistiky*, Praha (2011)
- [14] Lopatková, M., Kettnerová, V., Bejček, E., Vernerová, A., Žabokrtský, Z.: *Valenční slovník českých sloves VALLEX*. Karolinum, Praha (2016)
- [15] Kettnerová, V., Lopatková, M., Bejček, E.: The Syntax-Semantics Interface of Czech Verbs in the Valency Lexicon. In: *Proceedings of the XV EURALEX International Congress*, Oslo, University of Oslo (2012) 434–443
- [16] Lopatková, M., Kettnerová, V.: Alternations: From Lexicon to Grammar And Back Again. In Hajičová, E., Boguslavsky, I., eds.: *Proceedings of the Workshop on Grammar and Lexicon: Interactions and Interfaces (GramLex)*, Ōsaka, Japan, ICCL, The COLING 2016 Organizing Committee (2016) 18–27
- [17] Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., Suchomel, V.: The sketch engine: ten years on. *Lexicography ASIALEX* 1 (2014) 7–36
- [18] Kettnerová, V., Lopatková, M., Bejček, E., Vernerová, A., Podobová, M.: Corpus based identification of czech light verbs. In Gajdošová, K., Žáková, A., eds.: *Proceedings of the Seventh International Conference Slovko 2013; Natural Language Processing, Corpus Linguistics, E-learning*, Lüdenscheid, Germany, Slovak National Corpus, L' Štúr Institute of Linguistics, Slovak Academy of Sciences, RAM-Verlag (2013) 118–128