

The Representation of Czech Light Verb Constructions in a Valency Lexicon

Václava Kettnerová and Markéta Lopatková

Charles University in Prague

Faculty of Mathematics and Physics

Institute of Formal and Applied Linguistics

Czech Republic

{kettnerova, lopatkova}@ufal.mff.cuni.cz

Abstract

Light verb constructions (LVCs) pose a serious challenge for both theoretical and applied linguistics as their syntactic structures are not solely determined by verbs alone but also by predicative nouns. In this contribution, we introduce an initial step to a new formal lexicographic representation of LVCs for the valency lexicon of Czech verbs, *VALLEX*.

The main idea underlying our representation is to decompose the information on an LVC between (i) the verbal valency frame and (ii) the nominal valency frame. Both deep and surface syntactic structures of LVCs can be easily derived from the information given in the verbal and nominal frames by application of formal rules as they are introduced in this contribution.

1 Introduction

Light verb constructions (LVCs) represent a type of complex predicate where two syntactic elements serve as a single predicate – for example, in Czech, light verbs combine with predicative nouns, ex. (1), adjectives, ex. (2), or adverbs, ex. (3) (LVCs are typed in bold). These combinations are characterized by ambivalent relations: from a syntactic point of view, a light verb is the governing component of the collocation; however, from a semantic point of view, it is the predicative noun that represents the governing component.

- (1) *Petr získal **souhlas** od svého nadřízeného ke změně právního zástupce firmy.*
Eng. Peter **won approval** from his boss to change the legal representative of the company.
- (2) *Jan je **podobný** svému otci.*
Eng. John **is like** his father.

- (3) *Výpověď klíčového svědka by mohla **vnést** do případu **jasno**.*

Eng. Key witness testimony could **shed light** on the case.

Considering the wide range of issues, this study is limited to Czech LVCs based on the collocations of a light verb and a predicative noun, see ex. (1).

Despite being subject to many analyses, see esp. (Butt, 2010), a clear-cut definition of LVCs is still missing. In this paper, we follow both semantic and syntactic criteria for distinguishing LVCs. (i) The semantic operational criterion is based on the observation that a predicative noun (a predicative adjective or a predicative adverb) – as a semantic governing element – stands for the entire collocation of the predicative noun and a light verb; i.e. a predicative noun shows the same semantic distribution as the entire collocation. (ii) According to the syntactic criterion, some valency complementations of a light verb and a predicative noun have to be referentially identical, see esp. (Radimský, 2010) and (Kolářová, 2010).

Like other types of multiword expressions, LVCs pose a serious challenge for both theoretical and applied linguistics (Sag et al., 2002). As they require special treatment in NLP tasks, esp. in machine translation, their automatic recognition would bring a substantial benefit. Developing automatic recognition tools can be greatly assisted by lexical resources providing formal description of LVCs, see e.g. PropBank (Hwang et al., 2010), and WordNet (Vincze et al., 2012).

In this paper, we propose a formal representation of LVCs in the valency lexicon of Czech verbs, *VALLEX*.¹ The *VALLEX* lexicon is a collection of rich linguistically annotated data resulting from an attempt at a formal description of the valency behavior of Czech verbs. It provides the information on the valency structure of Czech

¹<http://ufal.mff.cuni.cz/vallex>

verbs in the form of valency frames, each valency frame corresponding to a single verbal lexical unit. For the description of valency, the valency theory formulated within the Functional Generative Description (FGD) – a dependency based framework – has been adopted (Sgall et al., 1986). Valency in FGD is related to the tectogrammatical layer – to a layer of linguistically structured meaning (roughly speaking, to a deep-syntactic layer). Five types of verbal actants – labeled by functors ACTor (ACT), PATient (PAT), ADDRessee (ADDR), ORIGin (ORIG), and EFFect (EFF) – have been determined. The first two – ACT and PAT – are distinguished on the syntactic basis. In assigning the remaining three actants – ADDR, ORIG and EFF – semantic criteria are taken into account as well, see esp. (Panevová, 1994).

The issue of LVCs (as well as other types of complex predicates) has remained underdeveloped in *VALLEX* so far. The main motivation of this paper is to propose an adequate representation of this phenomenon in order to fill such serious gap in the description of valency behavior of Czech verbs. When designing an LVCs representation for *VALLEX*, we draw inspiration esp. from the Explanatory Combinatorial Dictionary of the Contemporary Russian Language elaborated within the Meaning-Text Theory, in which a strong emphasis is put on the systematic description of combinatorial potentials of lexical units, see (Mel'čuk and Žolkovskij, 1984). For the description of LVCs, several lexical functions – allowing to identify verbonominal collocations – are used, see (Mel'čuk, 1996).

We have carried out a detailed analysis of semantic and deep syntactic aspects (Section 2) as well as surface syntactic aspects (Section 3) of the given constructions. The analysis of the LVCs based on the combination of light verbs with predicative nouns can be conducted either (i) from the perspective of a light verb (Gross, 1981), or (ii) from the perspective of a predicative noun (Mel'čuk, 1996). Each of these analyses poses different challenging problems. Moreover, they can lead to more or less different interpretations of forming complex syntactic structure of LVCs.

The representation of LVC proposed in this article combines these two perspectives: (i) From the *semantic point of view*, it is a predicative noun that provides the LVC with its semantic participants; thus semantic aspects of LVCs are described from

the perspective of a predicative noun here. (ii) On the other hand, *deep syntactic aspects* are described from the perspective of a (light) verb as it is a verb that provides its valency potential for semantic participants (evoked by the noun) and thus determines a core syntactic structure of a sentence.

The results of this analysis have been reflected in the representation of Czech LVCs in the *VALLEX* lexicon (Section 4). The proposed representation decomposes the information on LVCs between verbal and nominal lexicon entries, which are interlinked by a special attribute `-lvc`. Moreover, a special attribute `-map` attached to the verbal frame provides the information on the linking between verbal and nominal valency complementations referring to the same entities in LVCs. Based on this linking, deep and surface syntactic structure of LVCs can be derived by application of formal rules, which capture 'patterns' common to individual types of LVCs.

2 Semantic and (Deep) Syntactic Aspects

In this section, semantic and (deep) syntactic aspects of Czech LVCs are described in detail. When describing LVCs, it shows fruitful to distinguish:

- (i) semantic participants involved in the situation expressed by a given LVC (related to semantic content),² roughly corresponding to semantic actants in MTT, see (Mel'čuk, 2004a; Mel'čuk, 2004b),
- (ii) valency complementations (related to the deep syntactic layer), and
- (iii) surface syntactic positions (related to the surface syntactic layer).

Here the relation between semantic participants (Subsection 2.1) and valency complementations in LVCs is discussed (Subsection 2.2).

2.1 Semantic Participants

Verbonominal collocations forming LVCs represent a type of complex predicates where two syntactic elements – a light verb and a predicative noun – serve as a single predicate. In contrast to a single predicating verb, in LVCs, semantic features are decomposed between a light verb and a predicative noun.

²Generally, whereas the inventory of units of syntactic layers have been well elaborated, the inventory of semantic participants has not been satisfactorily compiled so far in FGD. Here we have adopted semantic roles used in FrameNet for the description of semantic participants.

As to the distribution of semantic properties, a light verb appears to be a semantically incomplete element expressing only general semantic properties (esp. aspectual nuances). To be semantically complete, it enters into the combination with a predicative noun which contributes individual lexical-semantic properties into the resulting complex predicate (Macháčková, 1979).

The following examples make evident that it is the noun (not the light verb) that determines the number of semantic participants (indicated by their semantic labels) expressed in LVCs, ex. (4)–(5), and their semantic features, ex. (6)–(7).

- (4) *Policista_{Speaker} podal hlášení o akci_{Inform} svému veliteli_{Recip}.*
 ‘The officer_{Speaker} – handed – a report – on the action_{Info} – to his commander_{Recip}.’
 Eng. The officer **reported** on the action to his commander.
- (5) *Sportovec_{Agent} podal v závodu velký výkon.*
 ‘The sportsman_{Agent} – handed – in the race – a great performance.’
 Eng. The sportsman **gave** a great **performance** in the race.
- (6) *Tento počítač / člověk_{in/animate} dělá hodně práce.*
 Eng. This computer/man_{in/animate} **does** much **work**.
- (7) *Tento *počítač / člověk_{in/animate} dělá velkou kariéru.*
 Eng. This *computer / man_{in/animate} **makes** a great **career**. (Radimský, 2010)

For instance, the light verb *obdržet* ‘to receive’ is depleted of individual semantic properties – including semantic participants, see ex. (9) – foregrounding only the abstract semantic facets (i.e., ‘transferring’) of its full verb counterpart; the latter expresses ‘transferring a physical object from an agent to a recipient’ characterized by three semantic participants, namely ‘Recipient’ (abbr. ‘Recip’), ‘Agent’, and ‘Theme’, see ex. (8).

- (8) *Výherce_{Recip} od nás_{Agent} obdrží drobný dárek_{Theme}.*
 Eng. ‘The winner_{Recip} **will receive** a small gift_{Theme} from us_{Agent}.’
- (9) *Velitel_{Recip} obdržel hlášení od policisty_{Speaker} o akci_{Inform}.*

Eng. The commander_{Recip} **received** the report on the action_{Inform} from the officer_{Speaker}.

To be semantically complete, the light verb combines with the predicative noun *hlášení* ‘report’ that denotes the situation of ‘conveying a piece of information to a recipient by a speaker’. This situation involves three participants – ‘Speaker’, ‘Recip’, and ‘Information’ (abbr. ‘Inform’). As a result, the situation expressed by the collocation *obdržet hlášení* ‘to receive report’ is the situation of reporting, characterized by the semantic participants provided by the predicative noun, see ex. (9).

2.2 Valency Complementations

From a (deep) syntactic point of view, both a predicative noun and a light verb in an LVC preserve their own valency potentials (represented in a form of valency frames), i.e., they are characterized each by own sets of valency complementations. In case of predicative nouns, valency frames represent the usage of nouns in nominal structures. In case of light verbs, we observe that in Czech valency frames are prototypically identical with the frames of their full verb counterparts. Thus we assume that the valency frames of light verbs are inherited from the valency frames of the respective full verbs, see Subsection 2.3.2. As light verbs – entering into combination with predicative nouns – form multiword lexical units, their valency frames describe some kind of ‘proto lexical units’ (in contrast to valency frames of full verbs, where valency frames correspond to lexical units).

For instance, both the verb *obdržet* ‘to receive’ and the noun *hlášení* ‘report’ forming the LVC *obdržet hlášení* ‘to receive report’ are characterized by their own valency frames: (i) The valency frame of the verb is inherited from the valency frame of the full verb (10), see ex. (11). (ii) The valency frame of the noun (12) represents the usage of the noun in a nominal structure, see ex. (13).

- (10) *obdržet* ‘to receive’ ... ACT PAT ORIG
 The valency complementations of the full verb are mapped onto the semantic participants ‘Recip’, ‘Theme’, and ‘Agent’, respectively; see ex. (11).
- (11) *Výherce_{ACT:Recip} od nás_{ORIG:Agent} obdrží drobný dárek_{PAT:Theme}.*
 Eng. ‘The winner_{ACT:Recip} **will receive** a small gift_{PAT:Theme} from us_{ORIG:Agent}.’

- (12) *hlášení* ‘report’ ... ACT ADDR PAT
The nominal valency complementations are mapped onto the semantic participants ‘Speaker’, ‘Recip’, and ‘Inform’ respectively, see ex. (13).

- (13) *Policistovo*_{ACT:Speaker} *hlášení*
*veliteli*_{ADDR:Recip} *o akci*_{PAT:Inform}
bylo stručné.
Eng. The officer’s_{ACT:Speaker} **report** on the action_{PAT:Inform} to his commander_{ADDR:Recip} was brief.

The correspondence between valency complementations of a verb and those of a noun in LVCs is discussed in the following subsections.

2.3 Linking of Verbal and Nominal Valency Complementations

A predicative noun (as was shown above) contributes its semantic participants (linked with nominal valency complementations) to the LVC. On the other hand, the verbal complementations are not semantically saturated, see Subsection 2.1. To acquire semantic capacity, the verbal complementations are interlinked with nominal ones (saturated by nominal semantic participants).

The linking of verbal complementations with the nominal ones is reflected in the notions of fusion or merger posited in connection with LVCs by authors from different theoretical backgrounds, see esp. (Alsina, 1997) and (Mohan, 1997). In FGD, this fact is tentatively referred to as sharing valency complementations between a verbal and a nominal valency frame which is indicated by a specific type of grammatical coreference – quasi-control, see esp. (Mikulová et al., 2006), (Kolářová, 2010), and (Cinková, 2009).

For instance, when the verb *obdržet* ‘to receive’ combines with the predicative noun *hlášení* ‘report’ into the LVC, the nominal valency complementations are still linked with nominal semantic participants (namely ‘Speaker’, ‘Recip’, and ‘Inform’, see (12) in Subsection 2.2). On the other hand, the verbal complementations do not correspond to any semantic participants. To acquire the semantic content, the verbal complementations are linked with the nominal complementations (and via them to the above given nominal semantic participants), see Fig. 1.

As to the mechanism of the linking: when the light verb *obdržet* ‘to receive’ combines with the

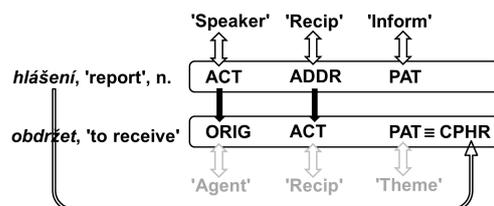


Figure 1: The linking of verbal valency complementations with nominal complementations (black arrows) and their saturation with the nominal semantic participants in the LVC *obdržet hlášení* ‘to receive report’.

predicative noun *hlášení* ‘report’, the noun occupies the verbal valency ‘PATient’. In accordance with FGD, we re-assign this valency complementation with the functor ‘CPHR’ (‘CompoundPHRASeme’) referring to a predicative component in complex predicates, see (Mikulová et al., 2006). The remaining valency complementations, ‘ORIGIn’ and ‘ACTor’ in the verbal frame (10), are linked with the nominal complementations ‘ACTor’ and ‘ADDRessee’ in (12), respectively. As argued above, the linking allows the given verbal complementations to acquire semantic capacity from the nominal complementations.

2.3.1 Direction of Linking

With respect to the fact that a change of a light verb may trigger the changes in the linking of verbal and nominal valency complementations, we assume that it is the light verb that determines the linking of its complementation(s).

For instance, according to our suggestion, the arrangement of the links is evoked by the verb *obdržet* ‘to receive’ (not by the noun *hlášení* ‘report’) in the LVC *obdržet hlášení* ‘to receive report’, see Fig. 1. This hypothesis is supported by the following observation: when the noun *hlášení* ‘report’ enters into combination with another light verb, e.g., with the verb *podat* ‘to hand’ (resulting in the LVC *podat hlášení* ‘to make report’), it leads to the rearrangement of the linking – in this case, the nominal ‘ACTor’ and ‘ADDRessee’ are linked with the ‘ACTor’ and ‘ADDRessee’ of the verb *podat* ‘to hand’, respectively, see ex. (14) and Fig. 2.

- (14) *Policista*_{ACT:Speaker} *podal hlášení*
*akci*_{PAT:Info} *svému veliteli*_{ADDR:Recip}.
Eng. The officer_{ACT:Speaker} **reported** on the action_{PAT:Info} to his

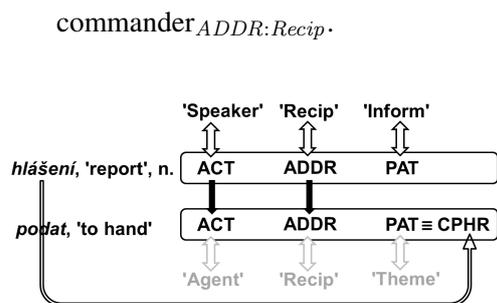


Figure 2: The linking of verbal valency complementations with nominal complementations (black arrows) and their saturation with the nominal semantic participants in the LVC *podat hlášení* ‘to make report’.

2.3.2 Verbal Valency Frame of a Light Verb

Let us repeat that from our point of view, valency frames of light verbs are inherited from the valency frames of their full counterparts, i.e., the valency frames of light verbs are prototypically identical with the frames of the respective full verbs (with the only difference that the complementation referring to a predicative noun is marked with ‘CPHR’ functor), as was stated in Subsections 2.2 and 2.3.

There is only one additional exception. In comparison with the valency frames of full verbs, the number of valency complementations in the frames of light verbs can be reduced. According to our proposal, only those verbal valency complementations from the valency frame that acquire semantic content from nominal ones are retained in the valency frame. Thus in case of light verbs, only those valency complementations (in addition to ‘CPHR’) that acquire semantic capacity via the linking with nominal complementations are employed in the valency frame. Those verbal complementations that are depleted in any semantic content (i.e., those that remain unlinked with any nominal complementation) are removed from the valency frame.

These cases occur when the number of nominal complementations is lower than the number of verbal ones left in the verbal valency frame after a predicative noun occupies some verbal complementation. Let us exemplify this case on the verb *podat* ‘to hand’ when it enters into combination with the predicative noun *výkon* ‘performance’ (resulting in the LVC *podat výkon* ‘to give performance’), see ex. (5). This predicative noun

is characterized by a single valency complementation – ‘ACTor’ corresponding to the semantic participant ‘Agent’. When this noun combines with the verb *podat* ‘to hand’, it fills the verbal ‘PATient’ (assigned with ‘CPHR’). Then two verbal complementations – ‘ACTor’ and ‘ADDRessee’ – remain left in the verbal frame. The verbal ‘ACTor’ is linked with the nominal ‘ACTor’; however, the verbal ‘ADDRessee’ remains unlinked. As a result, the ‘ADDRessee’ is deleted from the respective verbal valency frame, see Fig. 3.

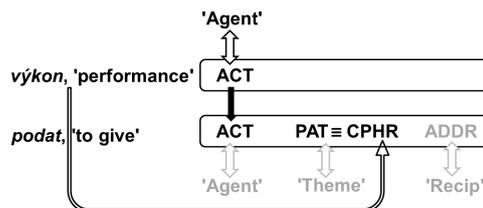


Figure 3: The linking of verbal valency complementations with nominal complementations (black arrows) and their saturation with the nominal semantic participants in the LVC *podat výkon* ‘to give performance’; see esp. the unlinked (and thus deleted) verbal ‘ADDRessee’ (in gray).

3 Syntactic Expressions and Morphemic Forms of Valency Complementations

We have argued that both a light verb and a predicative noun retain their own valency potentials, i.e., they correspond to separate valency frames, see Subsection 2.2. These valency structures enter into interaction which results in a complex surface syntactic structure of a LVC.

In general, the number of valency complementations that can be expressed on the surface is determined by the number of semantic participants involved in the situation expressed by an LVC, plus one verbal complementation (‘CPHR’) that is reserved for a predicative noun.

Czech, as an inflectional language encoding surface syntactic relations via morphological cases, gives us an excellent opportunity to study the role of valency complementations of a light verb and a predicative noun in the complex surface structure formation. According to morphemic forms of valency complementations expressed on the surface in LVCs, we can infer that the surface syntactic structure of an LVC is typically partly formed by a verbal valency frame and partly by a nominal valency frame.

We formulate the following hypothesis: It is a verb that, in general, determines the syntactic structure of a sentence. Thus in case that a particular semantic participant is linked with both a nominal valency complementation (directly) and a verbal valency complementation (via the link to a nominal valency complementation), it is a verb (not a noun) that retains the complementation in the resulted surface structure (and thus prescribe the morphemic form of the complementation). As each semantic participant is prototypically expressed only once, we consider the respective nominal complementation as elliptic (i.e., as having zero morphemic realization).

Based on this hypothesis, the following rules can be formulated:

- From the verbal valency frame, those verbal valency complementations that are semantically saturated via linking with some nominal valency complementations can be expressed on the surface.
- From the nominal frame, those nominal complementations that remain unlinked with any verbal ones are expressed on the surface. On the contrary, the nominal complementation affected by linking with verbal complementation remains unexpressed on the surface.³

Let us exemplify the proposed rules on the example of the LVC *obdržet hlášení* ‘to receive report’, which expresses the situation of reporting characterized by three situational participants – ‘Speaker’, ‘Recip’, and ‘Inform’ (as discussed in Section 2). Thus three valency complementations can be surface syntactically structured in the LVC (in addition to the valency complementation occupied by the predicative noun), see ex. (19).

As to the valency behavior, the light verb *obdržet* ‘to receive’ is characterized by the valency frame (16) inherited from the frame of its full counterpart (15). The valency frame of the predicative noun *hlášení* ‘report’ (17) describes the usage of the noun in a nominal structure, as in ex. (18).

³In some cases, ‘ACTor’ can be expressed twice on the surface, i.e., as both a nominal and a verbal complementation, despite being interlinked, e.g., *Nemocnice svůj boj proti rušení akutních lůžek nevzdávají*. ‘The **hospitals** do not give up **their** fight against eliminating acute beds.’ However, the possibility of expressing the ‘ACTor’ twice in a surface structure is subject to strong stylistic constraints in Czech.

- (15) *obdržet*: ACT_{nom} PAT_{acc} ORIG_{od+gen}
- (16) *obdržet*: ACT_{nom} CPHR_{acc} ORIG_{od+gen}
- (17) *hlášení*: ACT_{pos,gen} ADDR_{dat} PAT_{o+loc}
- (18) *Policistovo* ACT:pos *hlášení*
veliteli ADDR:dat *o akci* PAT:o+loc *bylo*
stručné.
 Eng. The officer’s_{ACT} **report** on the
 action to his commander_{ADDR} was brief.
- (19) *Velitel* ACT:nom *obdržel hlášení* CPHR:acc *o*
akci PAT:o+loc *od policisty* ORIG:od+gen.
 Eng. The commander_{ACT} **received re-**
port_{CPHR} on the action_{PAT} from the
 officer_{ORIG}.

1. When used in the LVC, one valency complementation of the verb *obdržet* ‘to receive’ – ‘PATient’ expressed in accusative case – is filled with the predicative noun *hlášení* ‘report’; instead of ‘PATient’, this complementation is marked by the ‘CPHR’ functor distinguishing the light verb from the full verb, see above.

2. Five valency complementations – two from the verbal frame (‘ACTor’ and ‘ORIGIN’) and three from the nominal one (‘ACTor’, ‘ADDRessee’, and ‘PATient’) – remain left in total for the expression of three semantic participants – ‘Speaker’, ‘Recip’, and ‘Inform’. Two verbal valency complementations ‘ACTor’ and ‘ORIGIN’ acquire semantic capacity from the nominal ‘ADDRessee’ and ‘ACTor’, respectively, see Subsection 2.2 and Fig. 1. Namely, the verbal ‘ACTor’ is linked with ‘Recip’ (via nominal ‘ADDRessee’) and the verbal ‘ORIGIN’ is linked with ‘Speaker’ (via nominal ‘ACTor’). According to our hypothesis, the verb retains these two complementations in the surface structure and it determines their morphemic forms, nominative and prepositional group *od+genitive*, respectively; see valency frame (15).

As a result, nominal ‘ADDRessee’ and ‘ACTor’, remain unexpressed on the surface, see Fig. 4 displaying the (simplified) dependency tree representing ex. (19) – the linked valency complementations are related by coreferential arrows going from the complementations unexpressed on the surface to the expressed ones.

3. The nominal ‘PATient’ – not being linked with any verbal complementation, see Fig. 1 – is expressed by the prepositional group *o+locative* modifying the noun as when the noun is used outside the LVC, see ex. (18) and (19).

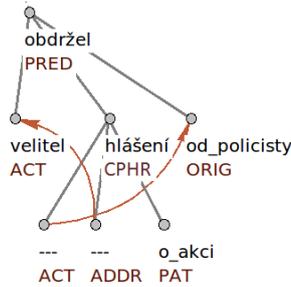


Figure 4: The (simplified) dependency tree for the LVC *obdržet hlášení* ‘to receive report’ in ex. (19).

The proposed hypotheses on the surface syntactic formation of LVCs deserve further examination on the corpus data.

3.1 Basic Typology of Syntactic Structures with Light Verb Constructions

We have identified three types of complex surface syntactic structures of LVCs in Czech (according to the linking criteria). They are briefly exemplified in the following paragraphs.

Type 1. All verbal complementations (excluding ‘CPHR’) are interlinked with the nominal ones and no nominal complementation remains unlinked. In this case, all the verbal complementations are expressed on the surface whereas all nominal complementations remain unexpressed. Compare with ex. (20) where the verbal ‘ACTor’ and ‘ADDRessee’ (linked with the nominal ‘ACTor’ and ‘ADDResse’, respectively) are realized in the resulting surface structure, whereas the respective nominal ones remain unexpressed (Fig. 5).

(20) *Janovi*_{ADDR} *poskytovala* *podporu*_{CPHR} *rodina*_{ACT}.

Eng. John’s family_{ACT} **provided support**_{CPHR} to him_{ADDR}.

Type 2. All verbal valency complementations (excluding ‘CPHR’) are linked with the nominal ones; however, some nominal complementations remain unlinked. In this case, the linked verbal complementations are expressed whereas the corresponding nominal complementations are unexpressed in the resulted surface structure (as in Type 1). Further, the unlinked nominal complementations remain expressed as nominal ones. Compare with ex. (21) where the verbal ‘ACTor’ (linked with the nominal ‘ACTor’) is expressed in the resulted surface structure, whereas the nominal ‘ADDRessee’ (being unlinked with any verbal

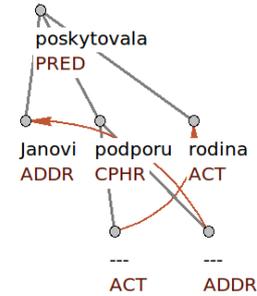


Figure 5: The (simplified) dependency tree for the LVC *poskytovat podporu* ‘to provide support’ in ex. (20).

complementation) is realized as a nominal complementation (expressed by the morphemic form determined by the given noun, Fig. 6).

(21) *Dceřin přítel*_{ACT} *na nás*_{ADDR} *udělal* *dojem*_{CPHR}.

Eng. The daughter’s boyfriend_{ACT} **made an impression**_{CPHR} on us_{ADDR}.

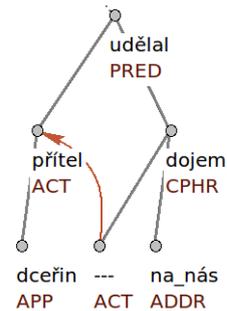


Figure 6: The (simplified) dependency tree for the LVC *udělat dojem* ‘to make impression’ in ex. (21).

Type 3. Not all verbal valency complementations (besides ‘CPHR’) are linked with the nominal ones, see Fig. 3. The linked ones are expressed on the surface whereas the unlinked ones are deleted from the verbal valency frame, see ex. (22) where the verbal ‘ADDRessee’ is not structured (Fig. 7), see also Subsection 2.3.2, Fig. 3.

(22) *Sportovec*_{ACT} *podal výkon*_{CPHR}.

Eng. The sportsman_{ACT} **gave a great performance**_{CPHR}.

4 Light Verb Constructions in VALLEX

In this section, the lexicographic representation of LVCs is proposed for the valency lexicon of Czech

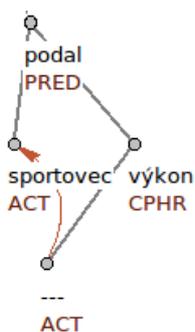


Figure 7: The (simplified) dependency tree for the LVC *podat výkon* ‘to give performance’ in ex. (22).

verbs, *VALLEX*. As every light verb and predicative noun creating an LVC are characterized by their own valency potentials, we represent them by separate valency frames: (i) for light verbs (Subsection 4.1) and (ii) for predicative nouns (Subsection 4.2). These frames are interlinked by references so that the whole collocations can be easily obtained. A special attention is paid to the representation of the mapping between valency complementations of predicative nouns and light verbs.

In the current version *VALLEX* 2.5, there are roughly 2,730 verb lexeme entries containing together around 6,460 verb lexical units; under the term lexical unit, we understand a form-meaning complex with (relatively) stable and discrete semantic properties. A lexeme then represents an abstract two-fold unit associating lexical form(s) with lexical unit(s). The verbs were selected according to their frequency in (the part of) the Czech National Corpus⁴ SYN2000 – the corpus coverage is approximately 98%. In building the lexicon, the main emphasis was laid on both human and machine readability – this is reflected in three formats of the lexicon: XML, HTML, and PDF formats.

The lexical entries of verbs in the *VALLEX* lexicon were exhaustively described in, e.g., (Žabokrtský and Lopatková, 2007). Let us shortly recapitulate here the basic information relevant for our explanation. Each lexical unit – represented by a lemma (or set of lemmas) – is characterized by obligatory attributes: gloss(es), a valency frame, and example(s). The valency frame, which provides the core information in the lexicon, is modeled as a sequence of valency slots. Each

slot stands for one valency complementation; it is characterized by a functor (indicating the type of the semantic relation of a valency complementation to a verb), by obligatoriness (in superscript), and by a list of possible morphemic form(s) (in subscript). In addition, optional attributes may follow (providing the information on syntactico-semantic class, the information of applicable alternations, etc.).

4.1 Representation of Light Verbs

In this subsection, we describe the necessary modification of verb lexical entries for the purpose of the representation of light verbs. Let us stress that light verbs in the *VALLEX* lexicon will be represented by ‘proto lexical units’ (proto-LUs) (as light verbs form multiword lexical units only in combination with predicative nouns, see Subsection 2.2). These proto-LUs are characterized by valency frames inherited from the frames of their full verb counterparts. Proto-LUs have to provide the following types of information:

I. In the inherited frame, the verbal valency complementation that is filled with a predicative noun is specified – its functor is changed to ‘CPHR’. In case that more verbal valency complementations can be filled by (different) predicative nouns, more inherited valency frames are determined, with different valency slots identified as ‘CPHR’. This functor covers the similar information which is captured by lexical functions (namely $Oper_i$, $Func_i$, and $Labor_{i,j}(k)$) in the Meaning-Text Theory, see esp. (Mel’čuk, 1996).

II. For each inherited valency frame, a list of possible linking(s) between the valency complementations of a light verb and those of a predicative noun is given in a special attribute *-map*, see Fig. 8.⁵ The following information can be drawn from the linkings:

- which of the given valency complementation(s) is/are expressed in a surface structure as verbal modification(s) (those complementations that are linked (via nominal complementations) with semantic participants, see Section 3, Type 1, 2 and 3), and
- which verbal valency complementation(s) is/are deleted from the verbal frame (those that are not linked) and thus cannot be expressed on the surface (Section 3, Type 3).

⁵As it is a light verb that forms the syntactic structure of a sentence, this information is listed within verbal frames, see Subsection 2.3.

⁴<http://ucnk.ff.cuni.cz>

III. Moreover, each inherited valency frame of a light verb contains the references to possible predicative nouns that form LVCs with the given light verb. As the mapping may differ for different predicative nouns, these references are attached to individual types of linkings.

For instance, the verb *podávat^{impf}*, *podat^{pf}* ‘to hand’ as a full verb is characterized by the lexical unit displayed on the top of Fig. 8. The valency frame representing the full verb is inherited by the proto-LU of the light verb; this proto-LU is characterized by the valency frame displayed below. In this valency frame, the ‘PATient’ is filled by predicative nouns (and thus replaced by the ‘CPHR’ functor). Moreover, two possible types of linking between verbal and nominal complementations are specified. Both linkings are attributed with the list of predicative nouns forming respective collocations.

<p>podávat^{impf}, podat^{pf} ‘to give, to hand’ LU1</p> <p>-gloss: <i>impf</i>: dávat do ruky <i>pf</i>: dát do ruky ‘to pass st to sb’s hand’</p> <p>-frame: ACT_{nom}^{obl} ADDR_{dat}^{obl} PAT_{acc}^{obl}</p> <p>-example: <i>impf</i>: podávat cihly bratrovi ‘to hand bricks to the brother’ <i>pf</i>: podat kolegovi šroubovák ‘to hand a screwdriver to the colleague’</p> <p>-class: exchange</p>
<p>podávat^{impf}, podat^{pf} ‘to give, to hand’ proto-LU1</p> <p>-frame: ACT_{nom}^{obl} ADDR_{dat}^{obl} CPHR_{acc}^{obl}</p> <p>-map₁: ACT_V – ACT_N; ADDR_V – ADDR_N</p> <p>-lvc₁: hlášení, námitka, návrh, oznámení, pokyn, stížnost, vysvětlení, zpráva, ...</p> <p>-map₂: ACT_V – ACT_N</p> <p>-lvc₂: demise, výkon, výpověď, ...</p>

Figure 8: The lexical unit of the full verb and the proto lexical unit of the light verb *podat*, *podávat* ‘to hand’ in *VALLEX*.

4.2 Representation of Predicative Nouns

The current version of the *VALLEX* lexicon covers only verbs, it does not comprise nouns. Thus for the purpose of the description of LVCs, it is necessary to enrich the lexicon with predicative nouns. The logical structure of *VALLEX* is designed in a way allowing for its further enriching with another part-of-speech.

As in case of verbs, each lexical unit of a noun is provided with a set of obligatory attributes providing the key information on the lexical unit – including a valency frame, gloss(es), and example(s). Again, the valency frame contains the core information on valency behavior of nouns. In case

of predicative nouns, each noun is assigned the valency frame corresponding to the usage of the noun outside LVC(s), see, e.g., the valency frame of the noun *hlášení* ‘report’ given in (17), Section 3 describing the usage of the noun in ex. (18).

In addition, a list of optional attributes that are applicable only to relevant lexical units may follow. Each predicative noun entering into combination with a light verb is attached with the optional attribute $-lvc$ containing references to possible light verbs forming LVCs with the given predicative noun;⁶ see, e.g., the proposed *VALLEX* lexical unit for the noun *hlášení* ‘report’ in Fig. 9.

<p>hlášení ‘report’</p> <p>-gloss: stručná zpráva, zvl. služební, úřední, vojenská apod. ‘a short information, esp. official announcement/statement’</p> <p>-frame: ACT_{pos,gen}^{obl} ADDR_{dat}^{obl} PAT_{o+loc}^{obl}</p> <p>-example: policistovo hlášení veliteli o průběhu; hlášení policisty veliteli o průběhu akce ‘the officer’s report on the course of the action to his commander’</p> <p>-lvc: verb: obdržet, podat, předat, přijmout, učinit, vypracovat, ...</p> <p>-class: reporting</p>

Figure 9: The proposed lexical unit in *VALLEX* describing the noun *podat* ‘to hand’.

5 Conclusion

We have provided a detailed analysis of the semantic and deep syntactic aspects of light verb constructions – we have explained the role of semantic participants with nominal and verbal valency complementations and their interlinking. We have also addressed the issue of surface syntactic expressions of LVCs by giving an explanation of changes in morphemic forms of the valency complementations affected in these constructions.

Our hypotheses have been projected to the proposal of the representation of Czech LVCs in the *VALLEX* lexicon. We have proposed to decompose the information on LVCs between (i) verbal valency frames (corresponding to light verbs) and (ii) nominal valency frames (corresponding to predicative nouns). Both frames are interlinked by special attribute $-lvc$.

The proposed representation reflects the close interplay between two components of LVCs – a

⁶The list of relevant light verbs are obtained automatically from the verbal part of the *VALLEX* lexicon – we suppose that for human readers, it is highly relevant to provide this information (also) within the nominal frame; the automatic extraction of such lists reduces duplicity (and thus decreases possible inconsistencies in the lexicon).

light verb and a predicative noun. The information provided by the -1_{VC} attribute assigned to prototypical units representing light verbs together with valency frames (inherited from full verbs) make it easy to derive both deep syntactic structure as well as surface syntactic structure and morphemic expressions of verbal and nominal valency complementations of LVCs.

As the future work, the proposed hypotheses on surface syntactic formation of LVCs will be further examined on the corpus data. Moreover, esp. three issues should be addressed in connection with LVCs: As for light verbs, a comprehensive inventory of their aspectual nuances should be compiled and included in their representation. As for predicative nouns, the restrictions imposed on the morphological category of number (e.g., *upadnout do rozpaků* ‘to fall into embarrassment’ where the Czech predicative noun can be used only in plural) deserve further theoretical research whose results should be covered in the lexicon as well. Further, the possibility of expressing some of valency complementations twice in the surface structure deserves further investigation.

Acknowledgments

This work has been using language resources stored and/or distributed by the LINDAT-Clarin project of MŠMT (project LM2010013). The research reported in this paper has been supported by GA ČR, grant No. GA P406/12/0557.

References

- Alex Alsina. 1997. Causatives in bantu and romance. In Alex Alsina, Joan Bresnan, and Peter Sells, editors, *Complex Predicates*, pages 203–246. CSLI Publications, Stanford, California.
- Miriam Butt. 2010. The Light Verb Jungle: Still Hacking Away. In Brett Baker Mengistu Amberber and Mark Harvey, editors, *Complex Predicates in Cross-Linguistic Perspective*, pages 48–78. Cambridge University Press, Cambridge.
- Silvie Cinková. 2009. *Words that Matter: Towards a Swedish-Czech Colligational Dictionary of Basic Verbs*, volume 2 of *Studies in Computational and Theoretical Linguistics*. UFAL, Prague.
- Maurice Gross. 1981. Les bases empiriques de la notion de prédicat sémantique. *Languages*, 63:7–53.
- Jena D. Hwang et al. 2010. Propbank annotation of multilingual light verb constructions. In *Proceedings of LAW 4*, pages 82–90, Uppsala, Sweden. ACL.
- Veronika Kolářová. 2010. *Valence deverbativních substantiv v češtině (na materiálu substantiv s dativní valencí)*. Univerzita Karlova, Nakladatelství Karolinum, Praha.
- Eva Macháčková. 1979. *Analytická spojení typu sloveso + abstraktní substantivum (analytické vyjadřování predikátů)*. Ústav pro jazyk český ČSAV, Praha.
- Igor A. Mel’čuk. 1996. Lexical Functions: A Tool for the description of lexical relations in a lexicon. In Leo Wanner, editor, *Lexical Functions in Lexicography and Natural Language Processing*, pages 37–105. John Benjamins, Amsterdam/Philadelphia.
- Igor A. Mel’čuk. 2004a. Actants in semantics and syntax I: actants in semantics. *Linguistics*, 42(1):1–66.
- Igor A. Mel’čuk. 2004b. Actants in semantics and syntax II: actants in syntax. *Linguistics*, 42(2):247–291.
- Igor A. Mel’čuk and Alexander Žolkovskij. 1984. *Tolkovo-kombinatornyj slovar’ sovremennovo russkovo jazyka*. Wiener Slawistischer Almanach, Sonderband 14, Wien.
- Marie Mikulová et al. 2006. Annotation on the teetogrammatical level in the Prague Dependency Treebank. Annotation manual. Technical Report TR-2006-30, ÚFAL MFF UK, Prague.
- Tara Mohanan. 1997. Multidimensionality of Representation: NV Complex Predicates in Hindi. In Alex Alsina et al., editor, *Complex Predicates*, pages 431–471, Stanford, California. CSLI Publications.
- Jarmila Panevová. 1994. Valency Frames and the Meaning of the Sentence. In Philip A. Luelsdorff, editor, *The Prague School of Structural and Functional Linguistics*, pages 223–243. John Benjamins Publishing Company, Amsterdam, Philadelphia.
- Jan Radimský. 2010. *Verbo-nominální predikát s kategoriálním slovesem*. Editio Universitatis Bohemiae Meridionalis, České Budějovice.
- Ivan A. Sag et al. 2002. Multiword Expressions: A Pain in the Neck for NLP. In *Proceedings of CLING 2002*, pages 1–15, Mexico City, Mexico.
- Petr Sgall, Eva Hajičová, and Jarmila Panevová. 1986. *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects*. Reidel, Dordrecht.
- Veronika Vincze, Attila Almási, and János Csirik. 2012. Multiword verbs in wordnets. In Ch. Fellbaum and P. Vossen, editors, *Proceedings of the 6th International Global WordNet Conference*, pages 377–381, Brno. Tribun.
- Zdeněk Žabokrtský and Markéta Lopatková. 2007. Valency information in VALLEX 2.0: Logical structure of the lexicon. *The Prague Bulletin of Mathematical Linguistics*, (87):41–60.