At the Lexicon-Grammar Interface: The Case of Complex Predicates in the Functional Generative Description

Václava Kettnerová and **Markéta Lopatková**
Charles University in Prague
Faculty of Mathematics and Physics
Czech Republic
{kettnerova,lopatkova}@ufal.mff.cuni.cz

Abstract

Complex predicates with light verbs have proven to be very challenging for syntactic theories, particularly due to the tricky distribution of valency complementations of light verbs and predicative nouns (or other predicative units) in their syntactic structure. We propose a theoretically adequate and economical representation of complex predicates with Czech light verbs based on a division of their description between the *lexicon* and the *grammar*. We demonstrate that a close interplay between these two components makes the analysis of the deep and surface syntactic structures of complex predicates reliable and efficient.

1 Introduction

Description of a language system is usually divided into two basic components – a grammar and a lexicon. The *grammar* consists of general patterns of a natural language rendered, in the form of formal rules which are applicable to whole classes of language units. The *lexicon*, on the other hand, represents an inventory of language units with their specific properties. Nevertheless, linguistic theories can substantially differ from each other in the distribution of information between the grammar and the lexicon.

Valency, which forms the core of a dependency structure of a sentence, constitutes a fundamental example of a phenomenon bridging between the grammar and the lexicon. Valency structure of verbs is so varied that it cannot be described by rules; it must be listed in lexical entries in a lexicon, see the highly elaborated lexicons, e.g., (Mel'čuk and Zholkovsky, 1984), (Apresjan, 2011). However, if a verb is a part of a complex predicate, its valency structure is involved in a complex structure the formation of which is typically regular enough to be described by rules in the grammar.

In this paper, we focus on lexicalized co-occurrence relations, namely on *complex predicates composed of light verbs and predicative nouns* (CPs) where two syntactic elements serve as a single predicate, e.g., 'to make a request', 'to give a presentation', 'to get support', 'to take a shower'.[1] We demonstrate that an adequate and economical description of CPs requires a close cooperation of the grammar and the lexicon: On the basis of the lexical representation of CPs, grammatical rules generate well-formed (both deep and surface) *dependency structures*.

The objective of this contribution is to further elaborate and modify – in light of recent investigations – the theoretical results given in (Kettnerová and Lopatková, 2013). Namely, the lexical information provided by the VALLEX lexicon (Lopatková et al., 2008) on diatheses and the grammatical rules in the grammatical component are applied to the description of CPs in *marked structures of diatheses* (e.g., passive structures) with the aim to gain all *surface syntactic manifestations* of the CPs.

The paper is structured as follows: first we discuss related work on CPs (Sect. 2); then we briefly introduce the Functional Generative Description (FGD) (Sgall et al., 1986) used as the theoretical background and the VALLEX lexicon (Sect. 3) and describe the lexical representation of CPs (Sect. 4); finally, we provide the enhancement of the grammatical component of FGD with formal rules for the generation of the syntactic structures with CPs (Sect. 5).

2 Related Work

There is a variety of approaches to *complex predicates with light verbs* (also called *light verb con-*

[1] Causative constructions of the type 'to make sb do something' are not considered here as CPs.

Proceedings of the Third International Conference on Dependency Linguistics (Depling 2015), pages 191–200,
Uppsala, Sweden, August 24–26 2015.

structions) and their characteristics, as well as to the range of issues involved in the notion of complex predicates. Despite the diversity in the treatment of complex predicates in different theoretical frameworks, there is a general agreement that the crucial issue to be resolved is that two syntactic elements function as a single predicate; this fact is corroborated by the presence of a single 'Agens'/'Bearer of action or property'/'Experiencer'. This key characteristic of complex predicates of the given type is accounted for by the mechanisms called argument fusion (Butt, 1998), argument transfer (Grimshaw and Mester, 1988), or argument composition (Hinrichs and Nakazawa, 1990) formulated within different theories.

All these mechanisms try to account for the fact that (i) light verbs, despite being depleted of semantic participants (denoting only general semantic scenario), have valency complementations, and that (ii) semantic participants (contributed to CPs primarily by predicative nouns) are usually expressed as complementations of light verbs (Alonso Ramos, 2007).

If a lexicographic representation aims at a description of syntactic behavior of CPs (not only at compiling an inventory of collocations of predicative nouns and light verbs, as e.g., (Vincze and Csirik, 2010), (Paul, 2010)), the above given mechanisms should be reflected in the lexicon. To our knowledge, the most complex representation of CPs is provided in the Explanatory Combinatorial Dictionary of Modern Russian (Mel'čuk and Zholkovsky, 1984) where the collocational potential is captured by means of lexical functions (Mel'čuk, 1996). The generation of well-formed syntactic structures with CPs is then based on the interplay of the lexical representation and grammatical rules (Alonso Ramos, 2007).

In Czech theoretical linguistics, there is only a limited number of studies devoted to CPs (Macháčková, 1994), (Cinková, 2009), (Radimský, 2010), and (Kolářová, 2010); none of them presents a mechanism aspiring to provide a thorough explanation of syntactic behavior of CPs. Moreover, the only existing lexical resource with information on syntactic properties of light verbs – PDT-Vallex – provides only partial information that does not make it possible to establish the deep and surface syntactic structures of the resulting CPs (Urešová, 2011).

3 FGD Framework

In this paper, we elaborate the representation of CPs within the Functional Generative Description, a stratificational and dependency-oriented theoretical framework (Sgall et al., 1986). One of the core concepts of FGD is that of valency (Panevová, 1994): at the layer of linguistically structured meaning (called the tectogrammatical layer), valency provides the structure of a dependency tree. The valency theory of FGD has been applied in several valency lexicons. The most elaborate one of these is the VALLEX, Valency Lexicon of Czech Verbs, which forms a solid basis for the lexical component of FGD.

VALLEX lexicon

The VALLEX lexicon[2] has resulted from an attempt to document valency behavior of Czech verbs (Lopatková et al., 2008). Over time, VALLEX has undergone many quantitative and qualitative extensions. Recent developments have focused on the linguistic phenomena that – despite representing productive grammatical processes involving changes in the valency structure of verbs – are lexically conditioned, esp. *diatheses*.

For the purposes of the representation of phenomena at the lexicon-grammar interface, VALLEX is divided into a lexical part and a grammatical part. The *lexical part* provides lexical representation of individual lexical units of verbs whereas the *grammatical part* represents formal representation of rules of the overall grammatical component of FGD that are directly connected to the valency structure of verbs.

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* of a verb, corresponding to its senses.

Each lexical entry of a verb is described by a set of attributes (see Fig. 2 below). The core attribute `frame` contains a valency frame that is modeled as a sequence of valency slots, each corresponding to a single valency complementation of the verb; each slot consists of (i) a functor – a syntactico-semantic label reflecting the type of dependency relation of the given valency complementation, (ii) an indication of obligatoriness, and (iii) a list of possible morphemic forms specifying the usage of a lexical unit in the *active voice*.

[2] `http://ufal.mff.cuni.cz/vallex`

Of all the remaining attributes of lexical units currently employed in VALLEX, we shall further discuss the attribute `diat`, the value of which is a list of all applicable diatheses (as their applicability is lexically conditioned and thus has to be captured in the *lexical part* of VALLEX). In the *grammatical part*, grammatical rules describing individual types of diatheses are formulated. When these rules are applied to the relevant lexical units (as indicated by the attribute `diat`), all possible surface syntactic manifestations of a lexical unit in the marked structures of diatheses can be obtained (Kettnerová et al., 2012).

4 Lexical Representation of CPs

A CP, as a multiword lexical unit, is formed as a combination of a predicative noun with an appropriate light verb. It is primarily the predicative noun that contributes its *semantic participants.* Its ability to select different light verbs (and thus to create different CPs) makes it possible to embed the event expressed by the predicative noun into different *general semantic scenarios* and thus to perspectivize it from the point of view of different semantic participants. In this process, a crucial role is played by the *referential identity* of nominal and verbal valency complementations within the CP (as it is demonstrated in Sect. 4.2.1).

As a consequence, CPs can be described as a combination of the information from the *valency frames* of both the *light verb* and the *predicative noun*. Further, we propose to enhance VALLEX with three special attributes `lvc`, `map` and `caus` to capture possible combinations of these two syntactic elements into a single predicate (Sect. 4.2).

4.1 Valency Frames

It is widely acknowledged that both predicative nouns and light verbs have their own valency potentials, i.e., they have their own sets of valency complementations (Alonso Ramos, 2007), (Macháčková, 1994). As a result, both light verbs and predicative nouns should be represented by their respective valency frames in the valency lexicon.

4.1.1 Predicative Nouns

Valency frames of predicative nouns underlie their deep dependency structures, both in nominal structures and as the nominal components of CPs, see examples (2) and (6) and the valency frame of the noun *pokyn* 'instruction' in (1).[3]

(1) ***pokyn***$_{\text{PN}}$ 'instruction':
$\text{ACT}_{\text{gen,pos}}$ ADDR_{dat} $\text{PAT}_{k+\text{dat,inf}}$

(2) ***Pokyn*** $_{\text{PN}}$ *státního zástupce* $_{\text{N:ACT:gen}}$ *žalobcům* $_{\text{N:ADDR:dat}}$ *(posuzovat případ jako krádež)* $_{\text{N:PAT:inf}}$ *přišel právě včas.*
'The **instruction**$_{\text{PN}}$ of the public prosecutor$_{\text{N:ACT}}$ to the prosecutors$_{\text{N:ADDR}}$ (to regard the case as a theft)$_{\text{N:PAT}}$ came just in time.'

Valency complementations of predicative nouns are endowed with semantic participants. For example, the noun *pokyn* 'instruction' is characterized by the participants 'Speaker', 'Recipient', and 'Information', which are mapped onto ACTor, ADDRessee, and PATient, respectively.

4.1.2 Light Verbs

Valency frames of light verbs constitute the deep dependency structure of the verbal component of CPs.

Formally, valency frames of Czech light verbs are prototypically identical to the valency frames of their full verb counterparts.[4] Hence we consider them to be inherited from the latter. The only regular difference between the valency frames of light verbs and their full verb counterparts is the functor CPHR 'Compound PHRaseme', indicating the valency position of the predicative noun.

Generally, valency complementations of a full verb correspond to its semantic participants; however, light verbs are deprived of semantic participants (Alonso Ramos, 2007).[5]

For example, the valency frame of the light verb *udělit*^{pf} 'to give, to grant' (4) is identical to the valency frame of the full verb (3), compare examples (5) and (6).

(3) ***udělit*** 'to give':
ACT_{nom} ADDR_{dat} PAT_{acc}

(4) ***udělit***$_{\text{LV}}$ 'to give':
ACT_{nom} ADDR_{dat} CPHR_{acc}

(5) *Prezident* $_{\text{V:ACT:nom}}$ ***udělil*** *umělcům* $_{\text{V:ADDR:dat}}$ *medaile* $_{\text{V:PAT:acc}}$.

[3] As the information on obligatoriness is not relevant here, we omit it from the valency frames.

[4] These findings are in line with the analysis of their morphological characteristics, which are also prototypically identical with the properties of their full counterparts (Butt, 2010).

[5] The only exception – causative light verbs – is addressed in Sect. 4.2.2.

'The President$_{V:ACT}$ **has awarded** medals$_{V:PAT}$ to the artists$_{V:ADDR}$.'

(6) *Státní zástupce* $_{V:ACT:nom}$ ***udělil*** $_{LV}$ *žalobcům* $_{V:ADDR:dat}$ ***pokyn*** $_{V:CPHR:acc}$ *posuzovat případ jako krádež.*
'The public prosecutor$_{V:ACT}$ **has given** an **instruction**$_{V:CPHR}$ to regard the case as a theft to the prosecutors$_{V:ADDR}$.'

Despite the absence of semantic participants of light verbs, their valency complementations are not semantically depleted: they acquire their semantic content from the semantic participants of predicative nouns via coreference with nominal valency complementations, as proposed, e.g., by (Butt, 1998), here Sect. 4.2.1. Then only semantically specified valency complementations are inherited from valency frames of full verb counterparts of light verbs (Kettnerová and Lopatková, 2013).[6]

4.1.3 Linking Valency Frames: Attribute `lvc`

For obtaining the deep dependency structure of a CP, the appropriate valency frames of the predicative noun and the light verb (with which the noun combines within the predicate) must be linked. In the VALLEX lexicon, the special attribute `lvc`, attached to individual valency frames of predicative nouns and (for convenience) also to those of light verbs, provides the list of references, see Fig. 1 and 2 below.

4.2 Lexical Mapping

The formation of well-formed deep and surface dependency structures with CPs requires a mechanism to account for the distribution of nominal and verbal valency complementations in the resulting syntactic structures. In this section, we show that for these purposes, additional information on the coreference of valency complementations (and thus on the mapping of semantic participants) has to be recorded as a part of lexical entries of predicative nouns and light verbs. This information is provided by two special attributes `map` (Sect. 4.2.1) and `caus` (Sect. 4.2.2).

[6] However, the cases in which the number of valency complementations in the valency frame of a light verb is reduced are rather rare in Czech (e.g., within the CP *přijmout zodpovědnost* 'to accept responsibility', the valency frame of the light verb does not inherit the ORIGin complementation as it lacks semantic specification).

4.2.1 Nominal Participants: Attribute `map`

As stated above, whereas the valency complementations of a predicative noun are semantically saturated by its semantic participants, the valency complementations of the light verb are semantically unspecified. To acquire semantic content, the verbal complementations enter in coreference relations with the nominal complementations. Pairs of nominal and verbal valency complementations within CPs thus exhibit referential identity (they refer to the same nominal semantic participant). This referential identity of verbal and nominal valency complementations represents a substantial characteristic of CPs.

For example, the CP *udělit pokyn* 'to give an instruction' can be characterized by three semantic participants given by the noun: 'Speaker', 'Recipient', and 'Information'. These participants are mapped onto the nominal valency complementations ACTor, ADDRessee, and PATient, see (1). The valency frame of the light verb in (4) comprises three complementations: one (CPHR) is occupied by the predicative noun and the remaining two (ACTor and ADDRessee) represent complementations that are not semantically specified by the light verb; however, they gain their semantic capacity via coreference with nominal ACTor and ADDRessee, see (7) specifying the referential identity.

(7) ***udělit pokyn*** 'to give an instruction':

$$\begin{array}{lll} \text{'Speaker'}_N & \Rightarrow \mathsf{ACT}_N & \Leftrightarrow \mathsf{ACT}_V \\ \text{'Recipient'}_N & \Rightarrow \mathsf{ADDR}_N & \Leftrightarrow \mathsf{ADDR}_V \\ \text{'Information'}_N & \Rightarrow \mathsf{PAT}_N & \end{array}$$

Due to the referential identity, all the valency complementations within this CP are semantically saturated. The event denoted by the predicative noun is perspectivized from the point of view of the 'Speaker', corresponding to the verbal ACTor (expressed in the active structure in the most prominent subject position, see also example (6).

Changes in the referential identity

The referential identity of the valency complementations may differ for different combinations of the same predicative noun combined with different light verbs (Kolářová, 2010), (Radimský, 2010).

For example, the referential identity within the CP *udělit pokyn* 'to give an instruction' (7) differs from that of the predicate *přijmout pokyn*

'to receive an instruction' (10). Within the latter, the same set of semantic participants are employed, i.e., 'Speaker', 'Recipient', and 'Information'. However, the verbal ACTor and ORIGin gain their semantic specification via coreference with the nominal ADDRessee and ACTor, respectively, see (1), (8) and (10).

(8) ***přijmout***$_{\text{LV}}$ 'to receive':
ACT$_{\text{nom}}$ CPHR$_{\text{acc}}$ ORIG$_{od+\text{gen}}$

(9) *Žalobci* $_{\text{V:ACT:}Recip}$ ***přijali*** $_{\text{LV}}$ *od státního zástupce* $_{\text{V:ORIG:}Speak}$ ***pokyn*** $_{\text{V:CPHR}}$ *(posuzovat případ jako krádež)* $_{\text{N:PAT:}Info}$.
'The prosecutors$_{\text{V:ACT:}Recip}$ **have received** the **instruction**$_{\text{V:CPHR}}$ (to regard the case as a theft)$_{\text{N:PAT:}Info}$ from the public prosecutor$_{\text{V:ORIG:}Speak}$.'

(10) ***přijmout pokyn*** 'to receive an instruction':
'Speaker'$_{\text{N}}$ $\Rightarrow$ ACT$_{\text{N}}$ $\Leftrightarrow$ ORIG$_{\text{V}}$
'Recipient'$_{\text{N}}$ $\Rightarrow$ ADDR$_{\text{N}}$ $\Leftrightarrow$ ACT$_{\text{V}}$
'Information'$_{\text{N}}$ $\Rightarrow$ PAT$_{\text{N}}$

The referential identity of valency complementations, provided in (10), reflects changes in the semantic specifications of verbal valency complementations (see example (9) illustrating the mapping) and also the change in the perspective from which the event expressed by the noun is viewed: in this case, the event is portrayed from the perspective of the 'Recipient' as the participant corresponding to the verbal ACTor.

Attribute `map`

As referential identity has a direct influence on the syntactic structure of CPs, see Section 5, this information has to be provided in the lexical part of the language description.

As it is the predicative noun that selects an appropriate light verb, the attribute `map` – giving a list of pair(s) of referentially identical nominal and verbal valency complementations – is assigned to *valency frames of predicative nouns*. More than one attribute `map` (distinguished by numeral indexes) can appear in a lexical unit of a predicative noun to account for the possible differences in referential identity of valency complementations within several CPs with the same predicative noun. Each attribute `map` is accompanied by a set of references to light verbs provided in the attribute `lvc` that comply with the given referential identity of valency complementations. The lexical entry is exemplified on the predicative noun *pokyn* 'instruction' in Fig. 1.

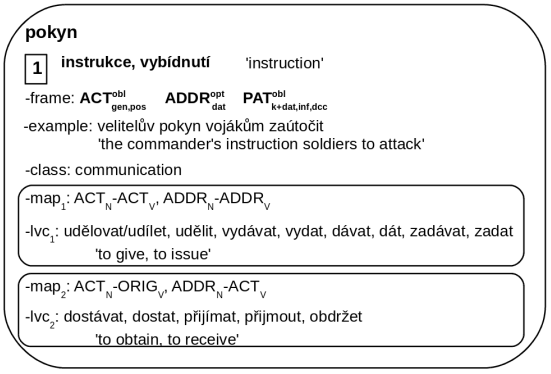


Figure 1: Simplified VALLEX lexical entry of the noun *pokyn* 'instruction'.

4.2.2 Verbal Participant 'Causator': Attribute `caus`

Typically, it is the predicative noun that determines the number and roles of semantic participants characteristic of a CP. Light verbs of causative type, which are endowed with the semantic participant 'Causator', represent the only exception. With these verbs, 'Causator' is contributed to CPs by the verb (in addition to the semantic participants provided by the predicative nouns).

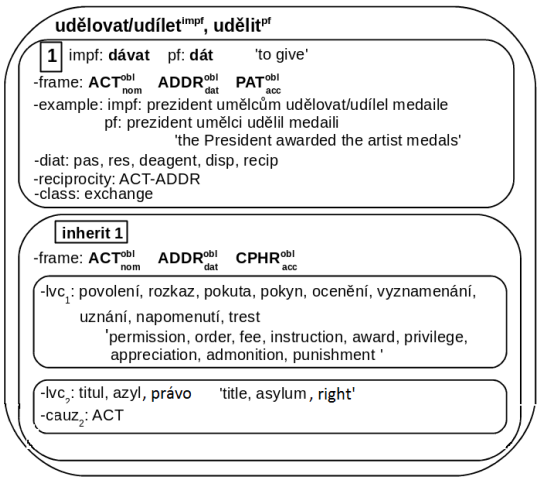


Figure 2: Simplified VALLEX lexical entry of the verb *udělovat/udílet*^{impf}, *udělit*^{pf} 'to give'.

For example, the CP *udělit právo* 'to grant a right', see example sentence (12), is characterized by three semantic participants: 'Causator', 'Bearer', and 'Theme'. 'Causator', provided by

the light verb *udělit* 'to grant' (with the valency frame given in (4)), is mapped onto the verbal ACTor whereas 'Bearer' and 'Theme' given by the predicative noun *právo* 'right' correspond to the nominal ACTor and PATient, respectively, see the valency frame of the noun in (11). As the verbal ACTor is saturated by the semantic participant 'Causator', only ADDRessee is not semantically saturated; this ADDRessee acquires its semantic specification from the predicative noun via coreference with the nominal ACTor, see their referential identity in (13). As a result, all valency complementations are semantically specified.

(11) ***právo***$_{\text{PN}}$ 'right':
ACT$_{\text{gen,pos}}$ PAT$_{\text{gen},na+\text{acc},\text{inf}}$

(12) *. . . král Vladislav Jagellonský* $_{\text{V:ACT:}Caus}$ ***udělil*** $_{\text{LV}}$ *městečku* $_{\text{V:ADDR:}Bearer}$ ***právo***$_{\text{V:CPHR}}$ *(pořádat dva výroční trhy)* $_{\text{N:PAT:}Theme}$.
'. . . king Ladislaus Jagiellon$_{\text{V:ACT:}Caus}$ **granted** the **right**$_{\text{V:CPHR}}$ (to hold two market fairs)$_{\text{N:PAT:}Theme}$ to the town$_{\text{V:ADDR:}Bearer}$.'

(13) ***udělit právo*** 'to grant a right':
'Causator'$_{\text{V}}$ $\Rightarrow$ ACT$_{\text{V}}$
'Bearer'$_{\text{N}}$ $\Rightarrow$ ACT$_{\text{N}}$ $\Leftrightarrow$ ADDR$_{\text{V}}$
'Name'$_{\text{N}}$ $\Rightarrow$ PAT$_{\text{N}}$

Changes in the mapping of 'Causator'

The semantic participant 'Causator' may be mapped not only onto the verbal ACTor but also onto another valency position of a light verb. Then the change in the mapping of 'Causator' brings about further changes in the referential identity of nominal and verbal complementations.

For example, within the CP *získat právo* 'to obtain a right', see (15), the 'Causator' contributed by the light verb *získat* 'to obtain' maps onto the verbal ORIGin, see the valency frame of this light verb in (14). In this case, it is the verbal ACTor that gains semantic content from the nominal ACTor (16). As a consequence, all the valency complementations within the CP *získat prá*vo 'to obtain a right' are semantically saturated.

(14) ***získat***$_{\text{LV}}$ 'to obtain':
ACT$_{\text{nom}}$ CPHR$_{\text{acc}}$ ORIG$_{od+\text{gen}}$

(15) *. . . od krále Vladislava Jagellonského* $_{\text{V:ORIG:}Caus}$ *městečko* $_{\text{N:ACT:}Bearer}$ ***získalo*** $_{\text{LV}}$ ***právo*** $_{\text{V:CPHR}}$ *(pořádat dva výroční trhy)* $_{\text{N:PAT:}Theme}$.
'. . . from king Ladislaus Jagiellon$_{\text{V:ORIG:}Caus}$, the town$_{\text{V:ACT:}Bearer}$ **obtained** the **right**$_{\text{CPHR}}$ (to hold two market fairs)$_{\text{N:PAT:}Theme}$.'

(16) ***získat právo*** 'to obtain a right':
'Causator'$_{\text{V}}$ $\Rightarrow$ ORIG$_{\text{V}}$
'Bearer'$_{\text{N}}$ $\Rightarrow$ ACT$_{\text{N}}$ $\Leftrightarrow$ ACT$_{\text{V}}$
'Name'$_{\text{N}}$ $\Rightarrow$ PAT$_{\text{N}}$

Attribute `caus`

The mapping of 'Causator' onto valency complementations is relevant for both deep and surface structure formation, therefore it is captured by a special attribute `caus` assigned to valency frames of light verbs of causative type. This attribute lists the verbal valency complementation onto which 'Causator' is mapped, see the light verb *udělovat/udílet*^{impf}, *udělit*^{pf} 'to give' in Fig. 2.

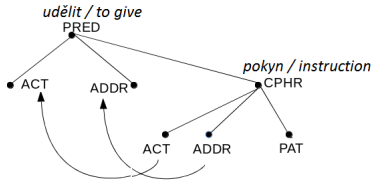
5 Grammatical Rules for CPs

The grammatical part of the VALLEX lexicon contains meta-rules describing the formation of deep (Sect. 5.1) and surface dependency structures of CPs (Sect. 5.2). These meta-rules are instantiated on the basis of the information stored in the lexical part of the lexicon.

5.1 Deep Syntactic Structure

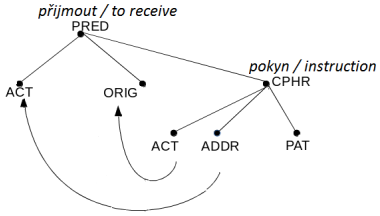
The meta-rule for formation of the deep syntactic structure of a CP exploits a valency frame of a predicative noun and a valency frame of a light verb with which the noun combines (their compatibility is identified by the attribute `lvc`). Moreover, information on the referential identity of nominal and verbal valency complementations within a CP, given in the attribute `map`, as well as information on verbal 'Causator', given in the attribute `caus` (if applicable), is necessary for the identification of coreferences in the dependency tree of the CP.

For example, the deep dependency structure of the CP *udělit pokyn* 'to give an instruction' is composed of the valency frame of the predicative noun *pokyn* 'instruction' and that of the light verb *udělit* 'to give' given above in (1) and (4), respectively. Further, the deep structure of this CP is characterized by coreferential links, reflecting the referential identity of the complementations, see (7), Fig. (17) (and Tab. 1 left part).



(17)

On the other hand, the valency structure of the CP *přijmout pokyn* 'to receive an instruction' results from the valency frames of the predicative noun *pokyn* 'instruction' and that of the light verb *přijmout* 'receive', given in (1) and (8), respectively, and from the referential identity provided in (10), see Fig. (18).



(18)

5.2 Surface Syntactic Structure

For the formation of the surface syntactic structure of a CP, its deep dependency structure is used (Sect. 5.1). In addition to the mapping of individual nominal and verbal complementations provided by the attribute `map` (Sect. 4.2.1), also the mapping of the verbal 'Causator', provided by the attribute `caus` (Sect. 4.2.2), is necessary.

Theoretical analysis has revealed that with CPs in Czech, each semantic participant is typically expressed in the surface sentence just once.[7] Despite the fact that semantic participants are contributed – with the exception of the verbal 'Causator' – by predicative nouns, Czech CPs have a strong tendency to express them in the surface structure as complementations of light verbs[8] (Macháčková, 1994). We propose the following rules for the formation of the surface syntactic structure with CPs:

All valency complementations from the *valency frame of the light verb* are expressed in the surface structure, namely:

(i) the valency complementation filled by the predicative noun (the CPHR functor);

(ii) the valency complementation corresponding to 'Causator' (the attribute `caus`);

(iii) valency complementations that are referentially identical with a nominal complementation (the attribute `map`).

Only the following valency complementations from the *valency frame of the predicative noun* are expressed in the surface structure:

(iv) valency complementations that are not referentially identical with any verbal complementation (i.e., those not listed in the attribute `map`).

For example, within the CP *udělit pokyn* 'to give an instruction' characterized by the deep dependency tree (17) the predicative noun fills the CPHR verbal position (i); two verbal valency complementations are expressed in the surface structure (iii), namely the ACT_V and ADDR_V (referentially identical with the ACT_N and ADDR_N, referring to 'Speaker' and 'Recipient', respectively); from the valency frame of the noun, only the PAT_N (referring to 'Information') is expressed on the surface (iv); the two remaining nominal complementations, ACT_N and ADDR_N, are unexpressed in the surface structure (as they are referentially identical with ACT_V and ADDR_V), see Tab. 1 column 4.

5.2.1 Unmarked (Active) Form

Morphemic forms of valency complementations of light verbs listed in the lexical part of the lexicon correspond to the *active form*. Thus the rules given above directly establish the surface syntactic structure of CPs in the active form.

For example, the surface structure of a sentence with the CP *udělit pokyn* 'to give an instruction' with the light verb in the active form can be obtained directly from morphemic forms recorded in the valency frames (1) and (4), see Tab. 1 column 5, and Fig. 3, displaying the surface syntactic tree of sentence (19) in relation to its deep dependency tree.

(19) *Státní zástupce* $_{\text{V:ACT:Sb}}$ ***udělil*** $_{\text{LV:active}}$ *žalobcům* $_{\text{V:ADDR:Obj}}$ ***pokyn*** $_{\text{V:CPHR:Obj}}$ *(posuzovat případ jako krádež)* $_{\text{N:PAT:Atr}}$.
'The public prosecutor$_{\text{V:ACT:Sb}}$ **has given** the prosecutors$_{\text{V:ADDR:Obj}}$ the **instruction**$_{\text{V:CPHR:Obj}}$ (to regard the case as a theft)$_{\text{N:PAT:Atr}}$.'

[7] The only exception is represented by the semantic participant mapped onto nominal ACTor; under certain conditions, this participant can be expressed twice, both as a verbal and as a nominal complementation (e.g., *Petr* $_{V:\text{ACT}:Bearer}$ *nevedl svůj* $_{N:\text{ACT}:Bearer}$ *život zrovna šťastně.* 'Peter did not lead his life very happily.').

[8] Rich morphology of Czech provides reliable clues for the identification of surface structure via morphemic cases.

CP	Deep	map & caus	Surface	active	pass	rcp-pass	deagent
Light verb	ACT_V		+	Sb:nom	Obj:instr, od+gen	Obj: od+gen	$-^*$
	$ADDR_V$		+	Obj:dat	Obj:dat	Sb:nom	Obj:dat
	$CPHR_V$		+	Obj:acc	Sb:nom	Obj:acc	Sb:nom
Predicat. noun	ACT_N	$ACT_N \Leftrightarrow ACT_V$	−				
	$ADDR_N$	$ADDR_N \Leftrightarrow ADDR_V$	−				
	PAT_N		+	Atr: k+dat,inf	Atr: k+dat,inf	Atr: k+dat,inf	Atr: k+dat,inf

Table 1: The deep (left part) and surface (right part) structures of the CP *udělit pokyn* 'to give an instruction'. (* The surface expression is blocked by the deagentive diathesis.)

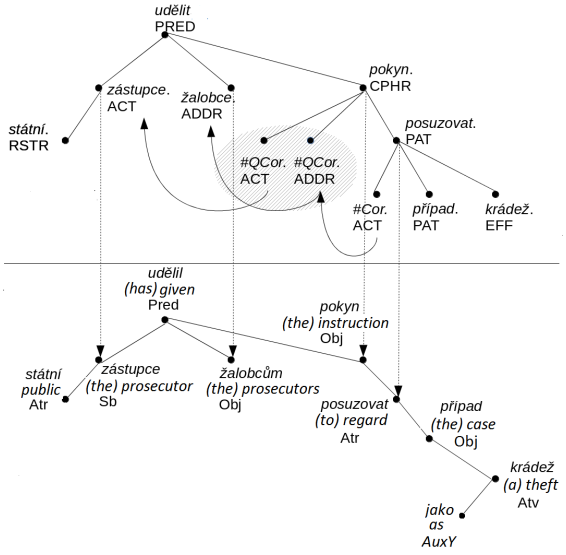


Figure 3: The simplified deep (above) and surface (below) dependency trees of sentence (19). The vertical arrows show the surface syntactic manifestations of valency complementations. The nominal valency complementations unexpressed in the surface structure (due to their referential identity with the verbal ones) are in the gray field.

5.2.2 Marked (Passive) Forms: Interplay of the Rules

The deep structure of a CP also serves as the basis for generating marked surface structures of diatheses. In this case, the rules for the formation of surface structures of CPs (Sect. 5.2 above) interplay with those for the formation of marked forms of diatheses (Vernerová et al., 2014).

In Czech, five types of diathesis (passive, resultative, recipient-passive, deagentive, and dispositional) were determined (Panevová et al., 2014). Diatheses are accompanied by changes in the morphological category of verbal voice and they are prototypically associated with shifts of valency complementations in the surface structure (while the deep structure is preserved). These shifts are indicated by changes in morphemic forms of the involved valency complementations and are regular enough to be captured by formal rules. These rules can be exemplified, e.g., by the rule for the recipient-passive diathesis:

`Rcp-pass d.`	
`verb form`	`replace(`active $\rightarrow$ $\rightarrow$ $AuxV_{dostat}$ + past_participle`)`
ACT	`replace(`nom $\rightarrow$ od+gen`)`
ADDR	`replace(`dat $\rightarrow$ nom`)`

The light verb and its full verb counterpart prototypically enter the same type of diatheses; the applicability of individual diatheses is provided by the attribute `diat` attached to the full verb.

For example, the light verb *udělit* 'to give, grant' can create the following marked structures (Fig. 2): passive (`pass` (20)), resultative (`res`), recipient-passive (`rcp-pass` (21)), deagentive (`deagent` (22)), and dispositional (`disp`) diathesis.

(20) *Žalobcům* V:ADDR:dat ***byl*** *od státního zástupce* V:ACT:od+gen ***udělen***$_{pass}$ ***pokyn*** V:CPHR:nom (*posuzovat případ jako krádež*) N:PAT:inf.
'The **instruction**$_{V:CPHR}$ (to regard the case as a theft)$_{N:PAT}$ **was given** to the prosecutors$_{V:ADDR}$ by the public prosecutor$_{V:ACT}$.'

(21) *Žalobci* V:ADDR:nom ***dostali*** *od státního zástupce* V:ACT:od+gen ***udělen***$_{rcp\text{-}pass}$ ***pokyn*** V:CPHR:acc (*posuzovat případ jako krádež*) N:PAT:inf.
'The prosecutors$_{V:ADDR}$ **have been given** the **instruction**$_{V:CPHR}$ (to regard the case as a theft)$_{N:PAT}$ by the public prosecutor$_{V:ACT}$.'

(22) *Žalobcům* V:ADDR:dat ***se udělil***$_{deagent}$ ***pokyn*** V:CPHR:nom (*posuzovat případ jako krádež*) N:PAT:inf.
'The **instruction**$_{V:CPHR}$ (to regard the case as a theft)$_{N:PAT}$ **was given** to the prosecutors$_{V:ADDR}$.'

Valency frames describing the marked structures of diatheses of a given CP can be generated on the basis of the rules for deriving the marked structures of diatheses (stored in the grammatical part of the VALLEX lexicon), applied to the *deep and surface active structures* of the CP. The deep dependency structure of the CP (i.e., the number and the type of its verbal and nominal valency complementations) is preserved whereas the surface syntactic expression of the verb and its complementations is changed as prescribed by the rule describing the respective diathesis (the surface form of the nominal valency complementations remains unchanged).

For example, the marked structure of the recipient-passive diathesis of the CP *udělit pokyn* 'to give an instruction', as in (21), is underlain by the valency frame obtained by the application of the above given rule to the valency frame corresponding to the active form of the light verb in (4), see Tab. 1 column 7.

6 Conclusion

In this paper, we have focused on complex predicates consisting of a light verb and a predicative noun. We have proposed their theoretically adequate and economical description based on the interplay between the grammatical and the lexical components of the language description. The special attributes `lvc`, `map` and `caus`, complying with the logical structure of the VALLEX lexicon as well as with the main tenets of the Functional Generative Description, were designed. The information provided in these attributes identifies recurrent patterns of light verb collocations (similarly as lexical functions into which it can be easily transferred), while grammatical rules in the grammatical component generate their well-formed (both deep and surface) dependency structures. We have shown how the proposed rules combine with the rules describing diatheses.

At present, a large-scaled lexicographic representation of light verbs is still missing despite the fact that these phenomena are widespread in the language (Kettnerová et al., 2013). We expect that the lexicon enriched with the information on light verbs will form a solid basis for their future integration into NLP applications which can substantially contribute to verifying the results of the proposed theoretical analysis.

Acknowledgments

The work on this project was supported by the grants of GAČR No. P406/12/0557 and GA15-09979S. This work has been using language resources distributed by the LINDAT/CLARIN project of the MŠMT No. LM2010013.

References

Margarita Alonso Ramos. 2007. Towards the Synthesis of Support Verb Constructions. In L. Wanner, editor, *Selected lexical and Grammatical issues in the Meaning–Text Theory*, pages 97–137. J. Benjamins, Amsterdam.

Valentina Apresjan. 2011. Active dictionary of the Russian language: Theory and practice. In I. Boguslavsky and L. Wanner, editors, *Meaning-Text Theory 2011*, pages 13–24, Barcelona. Universitat Pompeu Fabra.

Miriam Butt. 1998. Constraining argument merger through aspect. In E. Hinrichs, A. Kathol, and T. Nakazawa, editors, *Complex Predicates in Nonderivational Syntax, Syntax and Semantics*, pages 73–113. Academic Press, San Diego.

Miriam Butt. 2010. The Light Verb Jungle: Still Hacking Away. In M. Amberber, B. Baker, and M. 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*. Institute of Formal and Applied Linguistics, Prague.

Jane Grimshaw and Armin Mester. 1988. Light Verbs and Θ-Marking. *Linguistic Inquiry*, 19(2):205–232.

Erhard Hinrichs and Tsuneko Nakazawa. 1990. Subcategorization and VP structure in German. In S. Hughes and J. Salmons, editors, *Symposium on Germanic Linguistics*, Amsterdam. J. Benjamins.

Václava Kettnerová and Markéta Lopatková. 2013. The Representation of Czech Light Verb Constructions in a Valency Lexicon. In E. Hajičová, K. Gerdes, and L. Wanner, editors, *Proceedings of DepLing 2013*, pages 147–156, Praha. Matfyzpress.

Václava Kettnerová, Markéta Lopatková, and Eduard Bejček. 2012. The Syntax-Semantics Interface of Czech Verbs in the Valency Lexicon. In *Euralex International Congress 2012*, pages 434–443, Oslo. University of Oslo.

Václava Kettnerová, Markéta Lopatková, and Eduard Bejček et al. 2013. Corpus Based Identification of Czech Light Verbs. In K. Gajdošová and A. Žáková, editors, *Slovko 2013*, pages 118–128, Lüdenscheid. RAM-Verlag.

Veronika Kolářová. 2010. *Valence deverbativních substantiv v češtině (na materiálu substantiv s dativní valencí)*. Karolinum Press, Prague.

Markéta Lopatková, Zdeněk Žabokrtský, and Václava Kettnerová et al. 2008. *Valenční slovník českých sloves*. Karolinum Press, Prague.

Eva Macháčková. 1994. Constructions with Verbs and Abstract Nouns in Czech (Analytical Predicates). In S. Čmejrková and Fr. Štícha, editors, *The Syntax of Sentence and Text*, pages 365–375. J. Benjamins, Amsterdam.

Igor A. Mel'čuk and Alexander K. Zholkovsky. 1984. *Explanatory Combinatorial Dictionary of Modern Russian*. Wiener Slawistischer Almanach, Vienna.

Igor A. Mel'čuk. 1996. Lexical Functions: A Tool for the description of lexical relations in a lexicon. In L. Wanner, editor, *Lexical Functions in Lexicography and Natural Language Processing*, pages 37–105. J. Benjamins, Amsterdam.

Jarmila Panevová et al. 2014. *Mluvnice současné češtiny 2, Syntax na základě anotovaného korpusu*. Karolinum Press, Prague.

Jarmila Panevová. 1994. Valency Frames and the Meaning of the Sentence. In P. A. Luelsdorff, editor, *The Prague School of Structural and Functional Linguistics*, pages 223–243. J. Benjamins, Amsterdam.

Soma Paul. 2010. Representing Compound Verbs in Indo WordNet. In *GWC-2010*, Mumbai. The Global Wordnet Association.

Jan Radimský. 2010. *Verbo-nominální predikát s kategoriálním slovesem*. Editio Universitatis Bohemiae Meridionalis, České Budějovice.

Petr Sgall, Eva Hajičová, and Jarmila Panevová. 1986. *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects*. Reidel, Dordrecht.

Zdeňka Urešová. 2011. *Valence sloves v Pražském závislostním korpusu*. Ústav formální a aplikované lingvistiky, Prague.

Anna Vernerová, Václava Kettnerová, and Markéta Lopatková. 2014. To pay or to get paid: Enriching a valency lexicon with diatheses. In *LREC 2014*, pages 2452–2459, Reykjavík. ELRA.

Veronika Vincze and János Csirik. 2010. Hungarian Corpus of Light Verb Constructions. In *COLING 2010*, pages 1110–1118, Beijing.