Deverbal Nouns in Czech Light Verb Constructions

Václava Kettnerová, Veronika Kolářová, and Anna Vernerová

Charles University, Faculty of Matematics and Physics
Institute of Formal and Applied Linguistics
Malostranské nám. 25, 118 00, Prague 1, Czech Republic
`ketttnerova,kolarova,vernerova@ufal.mff.cuni.cz`

Abstract. In this paper, we provide a well-founded description of Czech deverbal nouns in both nominal and verbal structures (light verb constructions), based on a complex interaction between the lexicon and the grammar. We show that light verb constructions result from a regular syntactic operation. We introduce two interlinked valency lexicons, *NomVallex* and *VALLEX*, demonstrating how to minimize the size of lexicon entries while allowing for the generation of well-formed nominal and verbal structures of deverbal nouns.

Keywords: deverbal nouns, valency, light verb constructions, valency lexicon

1 Introduction

One of the long-standing interests of phraseology lies in the study of compositionality, i.e, to what extent the syntactic and semantic properties of multiword units can be deduced from the properties of their respective parts. Indeed, compositionality or the lack thereof has a wide range of consequences both in theoretical linguistics and in NLP applications (e.g., information retrieval and machine translation). In this paper, we focus on the syntactic formation of nominal structures of Czech deverbal nouns and verbal structures in which they take part, on light verb constructions representing multi-word units. We show that the deep and surface syntactic structures of light verb constructions can be treated as compositional, being composed by applying simple syntactic rules to the structure of deverbal nouns and the light verbs. Although we only investigate examples in Czech (taken mostly from the Prague Dependency Treebank), we believe the same mechanisms operate in other languages as well.

Czech deverbal nouns, being derived from verbs either by productive suffixes *-ní/tí*, e.g., *podporování* 'supporting', or by non-productive suffixes or the zero suffix, e.g., *podpora* 'support', can denote processes (in a broad sense, covering actions and states) as their base verbs (1), or abstract (2) or concrete (3) results of the processes denoted by their base verbs.

(1)   *podporování/podpora pacienta lékařem*
      'support(ing) of the patient by the doctor'

(2) *lékařova podpora pacientovi*
'doctor's support to the patient'

(3) *podpora v nezaměstnanosti*
'unemployment benefits'

Deverbal nouns are typically characterized by valency, i.e., by the ability to open a certain number of valency positions for other dependent units, representing their valency complementations. The description of valency structure of deverbal nouns is of great importance as deverbal nouns referring to processes (1) and to abstract results of processes denoted by their base verbs (2) – along with deadjectival nouns (e.g., *trpělivost* 'patience') or some primary nouns (e.g., *láska* 'love') – form not only nominal structures but verbal structures as well, the latter as parts of light verb constructions. In contrast to constructions with full verbs, the syntactic structure of light verb constructions is not determined solely by the verb but also by the deverbal noun with which the verb combines. Although the participation of light verbs and deverbal nouns in the syntactic structure formation of light verb constructions has attracted much attention of theoretical linguistics, see esp. [3], [1], [4], [13], many aspects of this process still remain unclear.

In this paper, we demonstrate a close interplay between valency structures of light verbs and deverbal nouns in the syntactic formation of Czech light verb constructions. Czech, an inflectional language encoding syntactic relations via morphological forms, provides an excellent basis for studying this mechanism because the forms clearly indicate whether a complementation is syntactically structured as a complementation of the light verb, or as a complementation of the deverbal noun. Our theoretical findings are applied in an extensive annotation of light verbs and deverbal nouns in the valency lexicons *VALLEX* and *NomVallex*, respectively. We show that the information provided by these lexicons is sufficient for generating well-formed light verb constructions.

The paper is structured as follows. First, we lay out the main tenets of the description of the valency structure of deverbal nouns (Section 2), with an emphasis on differences between the expression of valency complementations of deverbal nouns in nominal structures (Section 3) and in verbal structures (Section 4). Second, a lexicographic representation of light verbs and deverbal nouns is proposed (Section 5), which, supplemented with grammatical rules, allows for generation of well-formed light verb constructions.

2 Valency of Deverbal Nouns in FGD

In the description of the valency structure of deverbal nouns, we make use of the valency theory formulated within the Functional Generative Description (FGD), a dependency oriented framework which takes a stratificational approach to language description, see esp. [14]. In FGD, valency belongs to the tectogrammatical layer, i.e., a layer of linguistically structured meaning corresponding to the deep syntactic layer, see esp. [10]. The valency theory has been first elaborated for the

description of verbs. Corresponding to the argument-adjunct distinction in other linguistic theories, two kinds of valency complementations of verbs are distinguished in FGD – inner participants (actants) and free modifications (adjuncts) [12]. As actants, the valency complementations corresponding to the subject and the direct and indirect object are classified; their morphemic form is typically determined by the verb. Actants occur with a verb only once regardless of coordination and apposition (e.g., *(Jan a Marie)*$_\mathsf{ACT}$ *šli do kina.* '(John and Mary)$_\mathsf{ACT}$ went to the cinema.' provides only a single occurrence of the relevant complementation). Both obligatory[1] and optional actants characterize a verb in a unique way and in this sense, they have to be listed in its valency frame, see below. Five types of actants are distinguished mostly on syntactic criteria: 'Actor' (ACT), 'Patient' (PAT), 'Addressee' (ADDR), 'Effect' (EFF), and 'Origin' (ORIG) (e.g., *Vláda*$_\mathsf{ACT}$ *omezila těžbu*$_\mathsf{PAT}$ *uranu ze současných 950 tun*$_\mathsf{ORIG}$ *na 500 tun*$_\mathsf{EFF}$ *ročně.* 'The government$_\mathsf{ACT}$ restricted uranium mining$_\mathsf{PAT}$ from the current 950 tonnes$_\mathsf{ORIG}$ to 500 tonnes$_\mathsf{EFF}$ per year'). Among free modifications fall the valency complementations corresponding to adverbials; in contrast to actants, their morphemic form is not determined by the verb. Free modifications can typically occur more than once with a verb (e.g., *Až v neděli*$_\mathsf{TWHEN}$ *dopoledne*$_\mathsf{TWHEN}$ *3. září*$_\mathsf{TWHEN}$ *v 11 hodin*$_\mathsf{TWHEN}$ *oznámila britská vláda světu,...* 'Not before the morning$_\mathsf{TWHEN}$ on Sunday$_\mathsf{TWHEN}$ September$_\mathsf{TWHEN}$ the 3rd at 11 o'clock$_\mathsf{TWHEN}$ did the British government announce...'). Just as actants, free modifications are either obligatory, or optional. However, unlike actants, only obligatory free modifications characterize a verb in a unique way (e.g., *Petr přijel domů*$_\mathsf{DIR3}$. 'Peter arrived home$_\mathsf{DIR3}$.' and *Děti se dobře*$_\mathsf{MANN}$ *chovaly.* 'Children behaved well$_\mathsf{MANN}$.') and thus they have to be listed in the valency frame.

The same inventory of valency complementations is applied in the description of valency characteristics of deverbal nouns denoting processes and those that refer to abstract results of the processes. The inventory of valency complementations of deverbal nouns referring to concrete entities, especially to concrete results of processes, is broader, comprising other specific complementations such as the actant MATerial (e.g., *jedno balení másla*$_\mathsf{MAT}$ 'one package of butter$_\mathsf{MAT}$') and the free modifications APPurtenance (e.g., *oddělení odbytu*$_\mathsf{APP}$ 'sales department$_\mathsf{APP}$') and AUTHor (e.g., *výzdoba od Michelangela*$_\mathsf{AUTH}$ 'decoration by Michelangelo$_\mathsf{AUTH}$'), see [11].

The valency characteristics of both verbs and deverbal nouns are described in the form of valency frames. A valency frame is modeled as a sequence of valency slots. Each slot stands for one valency complementation and it comprises a functor (labeling the syntactic-semantic relation of the given complementation to its governing word) and the information on its obligatoriness; this information is supplemented with a list of possible morphemic forms which determine the expression of the complementation in the surface syntactic structure.

---

[1] We refer here to the obligatoriness on the tectogrammatical layer, i.e., on the deep syntactic layer; while being subject to different types of ellipsis on the surface layer, obligatory complementations are still present on the tectogrammatical layer.

3 Nominal Structures

Czech deverbal nouns express the meaning ranging from a process via an abstract result of the process to its concrete result. Those deverbal nouns that are derived from verbs by productive suffixes *-ní/-tí* typically preserve the meaning of the verbs from which they are derived; as a result, they mostly express processes as their base verbs (i). For example, the deverbal noun *plánování* expresses the process of planning just as its base verb *plánovat* 'to plan'. The deverbal nouns derived by non-productive suffixes or the zero suffix express processes only rarely. Typically, they either denote an abstract result of the process expressed by their base verbs (ii), or they refer to a concrete result of the process (iii). For example, the deverbal noun *plán* 'plan' derived by the zero suffix expresses both the abstract result of the planning process (e.g., *Janův plán studovat* 'John's plan to study') and its concrete result (e.g., *plán zahrady* 'a plan of the garden').

(i) The deverbal nouns referring to *processes* typically inherit the valency structure from their base verbs, i.e., the number and type of valency complementations and their obligatoriness remain the same. Only morphemic forms of those valency complementations that are expressed as the nominative subject and the accusative direct object of the base verb undergo systemic changes:

verb	$\rightarrow$ deverbal noun
nominative	$\rightarrow$ genitive, instrumental, possessive adjective or pronoun
accusative	$\rightarrow$ genitive, possessive adjective or pronoun

These changes are regular enough to be described in terms of rules [11]. For example, the valency frame of the deverbal noun *plánování* 'planning' denoting the process of planning, see (5), is identical with the valency frame of its base verb *plánovat* 'to plan', see (4), except for the morphemic forms of individual valency complementations. These morphemic forms exhibit systemic shifts; the form of the ACT is changed from nominative to one of the following forms: genitive, instrumental and a possessive adjective or pronoun, and the form of the PAT changes from accusative to genitive or a possessive adjective or pronoun while other forms remain unchanged.

(4) *plánovat* 'to plan': ACT^{obl}_{nom} $\mathsf{PAT}^{obl}_{acc,inf,dcc}$ [2]
   Firma ACT:nom *plánuje výstavbu* PAT:acc *elektrárny.*
   'The firm ACT is planning the construction PAT of a power plant.'

(5) *plánování* 'planning': $\mathsf{ACT}^{obl}_{gen,instr,poss}$ $\mathsf{PAT}^{obl}_{gen,poss,inf,dcc}$
   plánování výstavby PAT:gen *elektrárny firmou* ACT:instr
   'the firm's ACT planning of a power plant construction PAT'

(ii) The deverbal nouns denoting *abstract results* of the processes expressed by their base verbs behave similarly to the deverbal nouns expressing processes in that they typically exhibit the same valency structure as their base verbs in

---

[2] In the labels for morphemic forms, *nom, gen, dat, acc, loc* and *instr* stand for the cases, *inf* stands for an infinitive, *poss* stands for possessive adjectives and possessive noun forms, and *dcc* indicates the dependent content clause.

terms of the number and type of valency complementations (i.e., the functors and obligatoriness of the complementations). However, the morphemic forms of valency complementations of these deverbal nouns are – in addition to systemic shifts – often subject to non-systemic shifts as well [7]. For example, as in case of the deverbal noun *plánování* 'planning', the valency frame of the deverbal noun *plán* 'plan', see (6), referring to the abstract result of the process of planning, consists of the same number of valency complementations as the valency frame of its base verb *plánovat* 'to plan', see above (4). Further, their functors and their obligatoriness are preserved as well. However, in contrast to the deverbal noun *plánování* 'planning', the morphemic forms of the valency complementations of this deverbal noun undergo non-systemic shifts besides systemic ones, see the prepositional groups expressing the PAT *na*+accusative, *o*+locative and *pro*+accusative.

(6) *plán* 'plan': $\mathsf{ACT}^{obl}_{gen,poss}$ $\mathsf{PAT}^{obl}_{gen,poss,na+acc,o+loc,pro+acc,inf,dcc}$
plán firmy ACT:*gen* *na výstavbu* PAT:*na+acc* *elektrárny*
'plan of the firm ACT for the construction PAT of a power plant'

(iii) Valency frames of the deverbal nouns that refer to the *concrete results* of the processes denoted by their base verbs can substantially differ from the valency frames of the verbs. The changes in their valency frames can concern the number of valency complementations, their type with respect to both functors and obligatoriness, as well as possible morphemic forms. For example, the valency frame of the deverbal noun *plán* 'plan' with the concrete meaning (7) comprises an optional free modification AUTH (marked as typ) and an obligatory PAT.

(7) *plán* 'map': $\mathsf{AUTH}^{typ}_{gen,od+gen,poss}$ $\mathsf{PAT}^{obl}_{gen,poss}$
plán města PAT:*gen*
'map/plan of the city PAT'

4   Verbal Structures

In addition to nominal structures, the deverbal nouns denoting processes (type (i) in Section 3) and the deverbal nouns referring to abstract results of processes (type (ii) in Section 3) can be employed in verbal structures as well. In this case, the deverbal noun selects a particular light (semantically impoverished) verb to form a complex predicate. This complex predicate exhibits a discrepancy between semantic and syntactic behavior. It is the deverbal noun that represents the semantic core of the given predicate, contributing most if not all of the predicate's semantic participants. However, it is the light verb that syntactically governs the deverbal noun. In Czech, complex predicates of the given type form (idiomatic) light verb constructions.

Complex predicates in which the deverbal noun is expressed as the direct object of the light verb (i.e., by the prepositionless accusative) represent the central type of Czech complex predicates, e.g., *mít plán* 'to have a plan'. Rarely,

the deverbal noun can be expressed as the indirect object by a preposition-less case other than accusative, e.g., *dojít újmy$_{gen}$* 'to come to harm', *podrobit zkoumání$_{dat}$* 'to put under scrutiny', *zahrnout výčitkami$_{instr}$* 'to shower with reproaches', see esp. [9]. These predicates form light verb constructions (LVCs); the syntactic structure formation of LVCs is regular to a great extent (Section 4.1 and 4.2).

Less frequently, complex predicates in which the deverbal noun is expressed as an adverbial of the light verb occur in Czech, e.g., *mít v plánu* 'to intend, lit. to have in plan'. These predicates form the so-called idiomatic light verb constructions representing the borderline between LVCs and idioms due to a greater degree of irregularities in their syntactic formation (Section 4.4), see [2].

4.1 The deep syntactic structure

The deep structure of LVCs is formed by both valency complementations of the light verb and complementations of the deverbal noun within the complex predicate. For example, the deep structure of LVCs with the complex predicate *mít plán* 'have a plan' consists of both valency complementations of the deverbal noun *plán* 'plan', see the valency frame in (6) in Section 3, and complementations of the light verb *mít* 'have', see the frame in (9).

The valency frame of the deverbal noun corresponds to the usage of the noun in nominal structures (Section 3). Individual valency complementations of deverbal nouns are semantically saturated by semantic participants. For example, the ACT and PAT of the noun *plán* 'plan' are mapped onto 'Agent' and 'Goal', respectively.

The valency frame of the light verb is typically identical with the frame of its full verb counterpart, see the frame of the full verb *mít* 'have' in (8) and the frame of the light verb *mít* 'have' in (9) and examples (10) and (11), respectively. They only differ in that the valency position reserved for the deverbal noun, representing the nominal component of the complex predicate with the light verb, is marked by the functor CPHR.

(8)   *mít$_{full}$* 'have': ACT^{obl}_{nom} PAT^{obl}_{acc}

(9)   *mít$_{light}$* 'have': ACT^{obl}_{nom} CPHR^{obl}_{acc}

(10)   *Petr* ACT:nom *má pěkný dům* PAT:acc.
       'Peter ACT has a beautiful house PAT.'

(11)   *Petr* ACT:nom *má plán* CPHR:acc *přestavět dům.*
       'Peter ACT has a plan CPHR to rebuild the house.'

While valency complementations of full verbs are mapped onto semantic participants, complementations of light verbs do not correspond to any participants; the only exception is represented by light verbs with causative meaning that provide the semantic participant 'Causator', see below. For example, the ACT and PAT of the full verb *mít* 'have' are mapped onto 'Possessor' and 'Possession', respectively. In contrast, the ACT of the light verb *mít* 'have' (the only valency
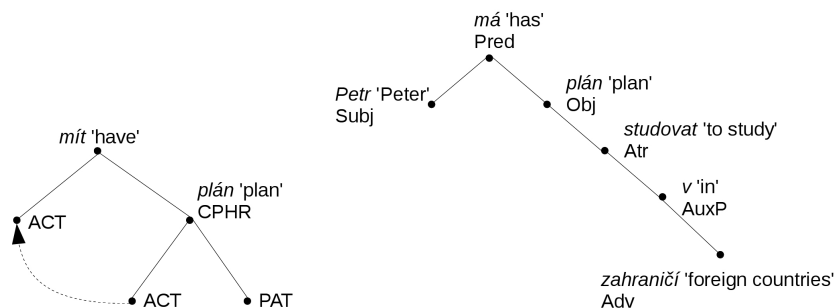
Fig. 1. The deep and surface syntactic structure of the complex predicate *mít plán* 'to have a plan'.

complementation in its valency frame except for CPHR) is not semantically saturated by any participant of the verb.

Although valency complementations of the light verb are not semantically specified by any semantic participants of the verb, they do not remain semantically unsaturated. To acquire semantic capacity, valency complementations of the light verb enter in coreference with complementations of the deverbal noun. For example, the ACT of the light verb *mít* 'have' within the complex predicate *mít plán* 'to have a plan' corefers with the ACT of the noun *plán* 'plan'; both these ACTors thus refer to the 'Agent', see the scheme of the mapping of the semantic participants onto valency complementations in the complex predicate *mít plán* 'to have a plan' provided in (12) ($\Rightarrow$ indicates the mapping of participants, $\rightarrow$ shows coreference). Figure 1 displays the deep syntactic structure of the complex predicate *mít plán* 'have a plan'.

(12) 'Agent'$_n$ $\Rightarrow$ ACT$_n$ $\rightarrow$ ACT$_v$
      'Goal'$_n$ $\Rightarrow$ PAT$_n$

In case of *causative light verbs*, the semantic participant 'Causator' is provided by the light verb, being mapped onto one of its valency complementations. For example, within the complex predicate *přinést radost* 'to bring joy' two semantic participants – 'Experiencer' and 'Stimulus' – are provided by the noun *radost* 'happiness, joy'. These participants correspond to the ACT and PAT of the noun, respectively, see the frame of the noun *radost* 'happiness, joy' in (13). In addition, the causative light verb *přinést* 'to bring' contributes the 'Causator', mapped onto its ACT, to the complex predicate *přinést radost* 'to bring joy', see the frame in (14). The ADDR of the light verb – the only remaining semantically unsaturated valency complementation in its valency frame – does not correspond to any participant. To be semantically saturated, the ADDR enters in coreference with the ACT of the deverbal noun *radost* 'happiness, joy', see the scheme of the mapping of the semantic participants within the complex predicate *přinést radost* 'to bring joy' in (15).

(13)   *radost* 'happiness, joy': $\mathsf{ACT}^{obl}_{gen,poss}$ $\mathsf{PAT}^{obl}_{nad+instr,z+gen,inf,dcc}$

(14)   *přinést* 'bring': ACT^{obl}_{nom} $\mathsf{ADDR}^{obl}_{dat}$ $\mathsf{CPHR}^{obl}_{acc}$

(15)   'Causator'$_v$ $\Rightarrow \mathsf{ACT}_v$
       'Experiencer'$_n$ $\Rightarrow \mathsf{ACT}_n \to \mathsf{ADDR}_v$
       'Stimulus'$_n$ $\Rightarrow \mathsf{PAT}_n$

4.2   The surface syntactic structure

The surface structure of LVCs can be composed of both valency complementations of the light verb and complementations of the deverbal noun. Each semantic participant is typically expressed in the surface structure only once. Despite being primarily contributed to LVCs by deverbal nouns, semantic participants tend to be expressed on the surface as valency complementations of light verbs. As corpus evidence shows, the surface structure formation of Czech LVCs is governed by the following principles [5]:

− From the *valency frame of the light verb*, all valency complementations are expressed, namely:
  - the valency complementation with the functor CPHR, indicating the deverbal noun,
  - the valency complementation corresponding to the 'Causator' (if present),
  - all valency complementations which corefer with complementations of the deverbal noun.

− From the *valency frame of the deverbal noun*, those valency complementations are expressed on the surface that are not in coreference with any complementation of the light verb.

For example, from the valency frame of the light verb *mít* 'to have', besides the CPHR reserved for the predicative noun, the ACT coreferring with the ACT of the deverbal noun *plán* 'plan' is expressed in the surface structure of the LVC with the complex predicate *mít plán* 'to have a plan'. From the valency frame of the noun *plán* 'plan', only the PAT, not coreferring with any complementation of the verb, is expressed on the surface; see the valency frames (6) and (9) and the scheme (12). Example (16) illustrates the LVC with the complex predicate *mít plán* 'to have a plan'; see also its surface syntactic structure in Figure 1.

(16)   *Petr*$_{v\mathsf{ACT}:nom}$ *má plán*$_{v\mathsf{CPHR}:acc}$ *studovat*$_{n\mathsf{PAT}:inf}$ *v zahraničí.*
       'Peter$_{v\mathsf{ACT}}$ has a plan$_{v\mathsf{CPHR}}$ to study$_{n\mathsf{PAT}}$ abroad.'

The surface structure of the LVC with the complex predicate *přinést radost* 'to bring joy' with the causative light verb *přinést* 'to bring' consists of all valency complementations of the light verb *přinést* 'to bring': the ACT corresponding to the 'Causator', the ADDR coreferring with the ACT of the deverbal noun *radost* 'happiness' (thus expressing the 'Experiencer') and the CPHR standing for the deverbal noun. From the valency frame of the deverbal noun, only the PAT that does not corefer with any complementation of the verb is expressed on the surface, see the valency frames (13) and (14), scheme (15) and example (17).

(17)   *Dětem*$_{v\mathsf{ADDR}:dat}$ *akce*$_{v\mathsf{ACT}:nom}$ *přinesla radost*$_{v\mathsf{CPHR}:acc}$ *z pěkných*
       dárků$_{n\mathsf{PAT}:z+gen}$.
     'The event$_{v\mathsf{ACT}}$ brought children$_{v\mathsf{ADDR}}$ joy$_{v\mathsf{CPHR}}$ from beautiful gifts$_{n\mathsf{PAT}}$.'

4.3   A quantitative comparison of nominal and verbal structures in corpus data

Prague Dependency Treebank (PDT),[3] containing Czech texts with complex and interlinked morphological, surface syntactic and deep syntactic annotation, provides an excellent basis for the study of syntactic behavior of deverbal nouns in nominal and verbal structures. We have divided the instances of nouns in PDT into two groups: the occurrences in which the noun has the CPHR functor,[4] representing verbal structures, and all other occurrences of nouns, representing their nominal structures. We have limited our investigation to the nominal lemmas that appear at least once in a verbal structure; Table 1 gives basic statistics of investigated lemmas and instances.

The graph in Figure 2 summarizes the frequency of different types of valency complementations of deverbal nouns expressed on the surface relative to all their occurrences; the remaining space to 100% is filled by the occurrences of that functor that are not expressed on the surface.

The figure shows that nominal ACT and ADDR are expressed much more rarely in verbal structures (where they typically corefer with verbal complementations) than in nominal structures. On the other hand, the PAT of productively derived nouns is expressed more often in verbal structures; a preliminary investigation of a sample of the *VALLEX* lexicon (Section 5.2) shows that a nominal PAT enters a coreference relation with a complementation of the light verb about three times less than a nominal ACT.

	lemmas	instances	
		nominal	verbal
productively derived (*-ní/tí*)	57	2265	159
non-productively derived	265	22028	1950
total	322	24293	2109

Table 1. Investigated PDT data

4.4   Remarks on Idiomatic Light Verb Constructions

By idiomatic light verb constructions (ILVCs) we understand the LVCs that are formed by the complex predicates within which the deverbal noun is expressed

---

[3] http://ufal.mff.cuni.cz/pdt3.0

[4] The instances of deverbal nouns governed by the verb *být* 'to be', representing a copula verb, were left aside.
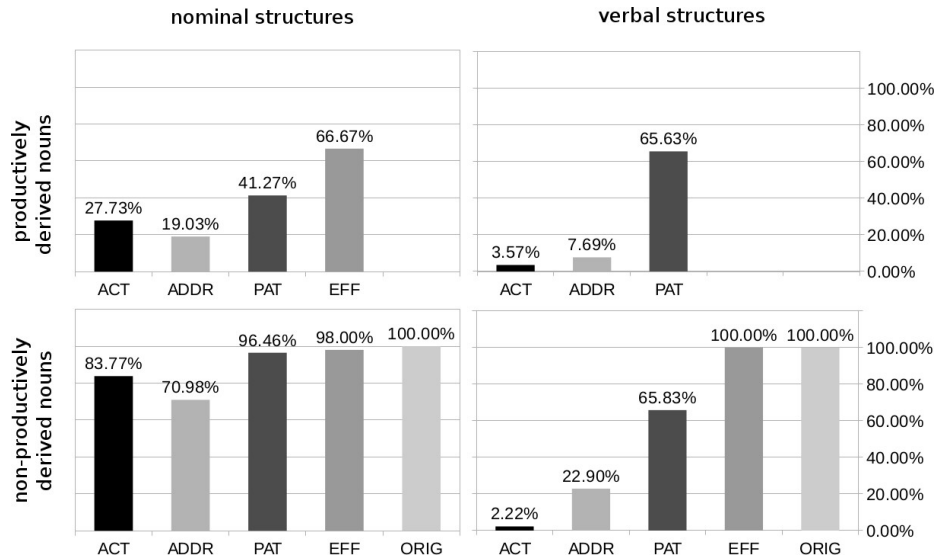
Fig. 2. For each functor, the percentage of its occurrences in which it is expressed in the surface structure of the sentence is given. The missing columns correspond to actants that did not appear with any noun instance in the investigated sample.

as an adverbial of the light verb (i.e., by various prepositional groups, e.g., *dát do pořádku* 'to put in order', *mít pod kontrolou* 'to have under control', *mít na starosti* 'to be responsible, lit. to have on care', *mít v úmyslu* 'to intend, lit. to have in intention').

In ILVCs, semantic participants are primarily provided – as in LVCs – by deverbal nouns that represent their semantic core. The only exception is represented by the complex predicates within which light verbs fulfill the causative function (e.g., *dostat pod kontrolu* 'to get under control', *uvést do chodu* 'to give a start, lit. lead into operation', *dát do ochrany* 'to give under protection, lit. to give in protection'), which contribute the participant 'Causator' initiating the event expressed by the deverbal noun.

The syntactic structure formation of ILVCs is governed by the same principles as that of LVCs:

- the *deep syntactic structure of ILVCs* results from an interplay between the valency frame of the deverbal noun and the frame of the light verb (Section 4.1) and it is characterized by coreference between valency complementations of the deverbal noun and complementations of the light verb;
- the *surface syntactic structure of ILVCs* is guided by the same principles as that of LVCs (Section 4.2).

For example, the deep structure of the ILVC formed by the complex predicate *mít v plánu* 'to intend, lit. to have in plan' is underlain by the valency frame of the light verb *mít* 'to have' (18) and the frame of the deverbal noun *plán*

'plan' (6) repeated here in (19). This predicate exhibits coreference between the nominal ACT and the verbal ACT and at the same time between the nominal PAT and the verbal PAT, expressing the 'Agent' and 'Goal' provided by the noun, respectively, see scheme (20) and the deep structure of this ILVC in Figure 3.

(18)   *mít* 'to have': ACT^{obl}_{nom} PAT^{obl}_{acc} $\text{CPHR}^{obl}_{v+loc}$

(19)   *plán* 'plan': $\text{ACT}^{obl}_{gen,poss}$ $\text{PAT}^{obl}_{gen,na+acc,o+loc,pro+acc,inf,dcc}$

(20)   'Agent'$_n$ $\Rightarrow$ $\text{ACT}_n \rightarrow \text{ACT}_v$
       'Goal'$_n$ $\Rightarrow$ $\text{PAT}_n \rightarrow \text{PAT}_v$

In the surface structure of the ILVC with the complex predicate *mít v plánu* 'to intend, lit. to have in plan', all valency complementations from the valency frame of the light verb (18) are expressed, namely ACT and PAT coreferring with the nominal ACT and PAT, and the CPHR occupied by the deverbal noun. No valency complementations from the valency frame of the deverbal noun, namely neither the ACT nor the PAT, are expressed on the surface as they corefer with the verbal complementations, see the principles of the surface structure formation in Section 4.2 and example (22). The surface structure of the ILVC is displayed in Figure 3.
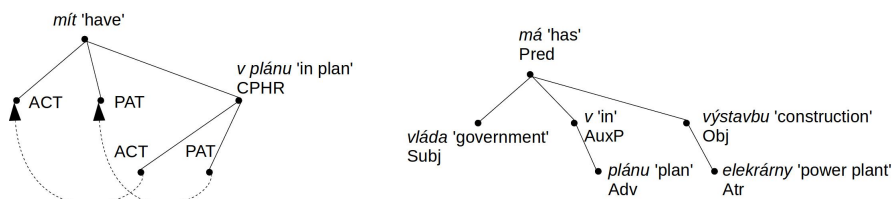

Fig. 3. The deep and surface syntactic structure of the complex predicate *mít v plánu* 'to have a plan, lit. to have in plan'.

However, ILVCs exhibit a lesser degree of regularity in morphemic expressions of valency complementations than LVCs. The irregularities concern morphemic forms of those valency complementations that semantically correspond to participants of propositional character. Corpus evidence shows that information on their morphemic forms cannot be easily inferred from the valency frame of the light verb containing the valency complementation referring to the given participant.

For example, the semantic participant 'Goal' characterizing the complex predicates *mít plán* 'to have a plan' and *mít v plánu* 'to intend, lit. to have in plan' represents a proposition. In LVCs formed by the complex predicate *mít plán* 'to have a plan', the 'Goal' is realized on the surface as an attribute of the deverbal noun *plán* 'plan', i.e., as its PAT (see principles in Section 4.2 and Figure 1). The PAT can be expressed in LVCs by any of the morphemic forms

prescribed in the valency frame of this noun, see the valency frame (19) and example (21).

In contrast, in ILVCs formed by the complex predicate *mít v plánu* 'to intend, lit. to have in plan', the Goal is expressed on the surface as the direct object of the light verb *mít* 'to have', i.e., as its PAT. In the valency frame of the light verb *mít* 'to have' (18) – underlying also this verb in the complex predicates without any propositional participant, e.g., *mít v užívání* 'to make use, lit. to have in use', *mít v oblibě* 'to have a liking, lit. to have in liking', only the form of a prepositionless accusative is prescribed for the given PAT. However, the PAT corresponding to the propositional participant 'Goal' in ILVCs with the complex predicate *mít v plánu* 'to intend, lit. to have in plan' can have the form of the infinitive and the dependent content clause as well, see (22).

(21) *Vláda*$_{v\text{ACT}}$ *má plán*$_{v\text{CPHR}}$ *výstavby*$_{n\text{PAT}:gen}$/*na výstavbu*$_{n\text{PAT}:na+acc}$/*o vý-stavbě*$_{n\text{PAT}:o+loc}$/*pro výstavbu elektrárny*$_{n\text{PAT}:pro+acc}$/*vystavět*$_{n\text{PAT}:inf}$/ *že vystaví*$_{n\text{PAT}:dcc}$ *elektrárnu.*

(22) *Vláda*$_{v\text{ACT}}$ *má v plánu*$_{v\text{CPHR}}$ *výstavbu*$_{v\text{PAT}:acc}$/*vystavět*$_{v\text{PAT}:inf}$/ *že vystaví*$_{v\text{PAT}:dcc}$ *elektrárnu.* 'The government$_{v\text{ACT}}$ has intention$_{v\text{CPHR}}$ to construct$_{v\text{PAT}}$ a power plant.'

To capture irregularities in the expression of propositional participants in ILVCs, we establish the following principle:

– The valency complementation of the light verb which in the surface structure of ILVCs occupies the direct object may be expressed, additionally to the forms listed in the valency frame of the light verb, by an infinitive or a dependent content clause on condition that these forms are listed as a morphemic expression of the coreferring nominal complementation.

5 Lexicographic Representation of Deverbal Nouns

In this section, we introduce *NomVallex* and *VALLEX*, valency lexicons which provide information on the basis of which nominal (Section 5.1) and verbal structures containing deverbal nouns (Section 5.2) can be generated.

5.1 Nominal Structures in a Lexicon

The valency lexicon *NomVallex* provides all the information necessary for producing nominal structures governed by Czech deverbal nouns, together with additional syntactic information [6]. This lexicon has adopted the design of the valency lexicon of Czech verbs, *VALLEX*, see Section 5.2.

In *NomVallex*, each deverbal noun is described by a lexeme, an abstract twofold unit associating all lexical forms of the noun with lexical units (individual senses of the noun). Each nominal lexeme is represented by a set of lemma(s); it is formed by a set of lexical unit(s) corresponding to the individual senses. The

lexical units are specified with respect to their types, i.e., whether they denote processes, abstract results or concrete results of processes.

Key information on the valency structure of each lexical unit is provided by the valency frame, see Section 2. Besides the valency frame, each lexical unit is accompanied with a gloss describing the given meaning and an example providing its corpus evidence. Further, each lexical unit can be assigned additional syntactic and semantic information, e.g., on reciprocity, control, and semantic class membership.

Each deverbal noun forming complex predicates with light verbs has the attribute `lvc` providing references to individual valency frames of light verbs with which the noun combines; the particular light verbs are stored in the *VALLEX* lexicon, see Section 5.2. On the basis of these references, individual complex predicates can be obtained. See Figure 4 displaying a simplified entry of the deverbal noun *plán* 'plan'.

* PLÁN

: id: blu-n-plán-1   plán
+ $\mathsf{ACT}^{obl}_{gen,poss}$ $\mathsf{PAT}^{obl}_{gen,na+acc,o+loc,pro+acc,inf,aby,že,dcc}$
   -derived: blu-v-plánovat-1
   -gloss: *záměr, úmysl* 'intention, aim'
   -example: *plány vystavět*$_{\mathsf{PAT}}$ *podnik* 'plans to construct a processing plant'
   -lvc: blu-l-mít-2, . . .
   -control: ACT, ex
   -class: mental action
   -type: abstract result

: id: blu-n-plán-2   plán
+ $\mathsf{AUTH}^{typ}_{gen,poss}$ $\mathsf{PAT}^{obl}_{gen,poss}$
   -derived: blu-v-plánovat-1
   -gloss: *mapa, nákres* 'map, layout'
   -example: *plán města*$_{\mathsf{PAT}}$ 'city plan'
   -class: mental action
   -type: concrete

Fig. 4. Simplified entry of the noun *plán* 'plan'.

5.2 Verbal Structures in a Lexicon

The derivation of verbal structures with deverbal nouns (LVCs and ILVCs, see Section 4) requires a close cooperation between the lexicographic representation of deverbal nouns (see Section 5.1) and light verbs on the one hand, and grammatical rules, on the other (see Section 4). The representation of LVCs is proposed here for the valency lexicon of Czech verbs *VALLEX*, see [8]. The structure of this lexicon is the same as that of *NomVallex* (Section 5.1).

Light verbs in *VALLEX* are treated as specific senses of verbs. Each light verb is assigned its respective valency frame. The valency position reserved in the valency frame for deverbal nouns is labeled by the CPHR functor. Each valency frame of the light verb is assigned three special attributes: `lvc`, `instig`, and `map`. The attribute `map` provides a list of pairs of valency complementations of the deverbal noun and the light verb that are in coreference. The attribute `lvc` provides references to relevant deverbal nouns. If relevant, the attribute `instig` introduces that valency complementation of the light verb onto which the semantic participant 'Causator' is mapped. See Figure 5 displaying the simplified entry of the verb *mít* 'to have'.

The lexicon thus economically captures recurring patterns of complex predicates with the given light verb, and the grammatical rules enable users to generate well-formed LVCs and ILVCs.

* MÍT

: id: blu-v-mít-1   **impf:** mít **iter:** mívat
+ ACT^{obl}_{nom} PAT^{obl}_{acc} LOC^{typ}
   -`gloss`: vlastnit 'to possess'
   -`example`: *Petr má dům v Karpatech.* 'Peter has a house in Carpathians.'

: id: blu-l-mít-2   **impf:** mít **iter:** mívat
+ ACT^{obl}_{nom} $CPHR^{obl}_{acc}$
   -`lvc`: blu-n-plán-1, ...
   -`map`: ACT_v–ACT_n
   -`example`: *Petr má plán studovat v zahraničí.* 'Peter has a plan to study abroad.'

: id: blu-l-mít-3   **impf:** mít **iter:** mívat
+ ACT^{obl}_{nom} PAT^{obl}_{acc} $CPHR^{obl}_{v+loc}$
   -`lvc`: blu-n-plán-1, ...
   -`map`: ACT_v–ACT_n, PAT_v–PAT_n
   -`example`: *Petr má v plánu studovat v zahraničí.* 'Peter has a plan to study abroad.'

Fig. 5. Simplified entry of the verb *mít* 'to have', showing the valency frames for the core meaning of the full verb and its corresponding light verbs.

6   Conclusion

In this paper, we have proposed a lexicographic representation of nominal and verbal structures of Czech deverbal nouns, focusing primarily on the verbal ones, i.e., light verb constructions and idiomatic light verb constructions. We have demonstrated that their well-formed syntactic structure can be derived by a complex process requiring a close interplay of information provided by lexicons and grammatical rules, making it possible to better understand the syntactic compositionality of these types of multi-word units.

Acknowledgements

References

1. Alsina, A., Bresnan, J., Sells, P. (eds.): Complex Predicates. CSLI Publications, Stanford, California (1997)
2. Baldwin, T., Kim, S.N.: Multiword expressions. In: Indurkhya, N., Damerau, F.J. (eds.) Handbook of Natural Language Processing, pp. 267–292. CRC Press, Boca Raton, USA (2010)
3. Grimshaw, J., Mester, A.: Light verbs and θ-marking. Linguistic Inquiry 19(2), 205–232 (1988)
4. Hinrichs, E., Kathol, A., Nakazawa, T.: Complex Predicates in Nonderivational Syntax. Academic Press, San Diego (1998)
5. 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 University (2015)
6. Klímová, J., Kolářová, V., Vernerová, A.: Towards a corpus-based valency lexicon of Czech nouns. In: GLOBALEX 2016: Lexicographic Resources for Human Language Technology. pp. 1–7 (2016)
7. Kolářová, V.: Special valency behavior of Czech deverbal nouns, chap. 2. John Benjamins, Amsterdam – Philadelphia (2014)
8. Lopatková, M., Kettnerová, V., Bejček, E., Vernerová, A., Žabokrtský, Z.: VALLEX 3.0 - Valenční slovník českých sloves. Karolinum, Praha (2016), `http://ufal.mff.cuni.cz/vallex/3.0/`
9. Macháčková, E.: Constructions with verbs and abstract nouns in Czech (analytical predicates). In: Čmejrková, S., Štícha, F. (eds.) The Syntax of Sentence and Text, pp. 365–374. John Benjamins, Amsterdam, Philadelphia (1994)
10. Panevová, J.: Valency frames and the meaning of the sentence. In: Luelsdorff, P.A. (ed.) The Prague School of Structural and Functional Linguistics, pp. 223–243. John Benjamins, Amsterdam – Philadelphia (1994)
11. Panevová, J.: K valenci substantiv (s ohledem na jejich derivaci). In: Zbornik Matice srpske za slavistiku, pp. 29–36 (2002)
12. Panevová, J.: In favour of the argument-adjunct distinction (from the perspective of FGD). The Prague Bulletin of Mathematical Linguistics (106), 21–30 (2016)
13. Ramos, M.A.: Towards the synthesis of support verb constructions: Distribution of syntactic actants between the verb and the noun. In: Wanner, L., Meľčuk, I.A. (eds.) Selected Lexical and Grammatical Issues in the Meaning-Text Theory, pp. 97–137. John Benjamins, Amsterdam – Philadelphia (2007)
14. Sgall, P., Hajičová, E., Panevová, J.: The Meaning of the Sentence in Its Semantic and Pragmatic Aspects. D. Reidel Publishing Company, Dordrecht (1986)