Representation of Changes in Valency Structure of Verbs in the Valency Lexicon of Czech Verbs[*]

Václava Kettnerová and **Markéta Lopatková**
Charles University in Prague
Institute of Formal and Applied Linguistics
{kettnerova,lopatkova}@ufal.mff.cuni.cz

1 Introduction

Valency behavior of verbs is so multifarious that it cannot be described by formal rules; instead, it must be listed in the form of a lexical entry separately for each verb. Prototypically, a single verbal meaning corresponds to a single valency frame. However, in many cases, semantically close uses of verbs can be syntactically structured in different ways. See the following examples:

(1) a. *Petr namazal máslo na chleba.*
 b. *Petr namazal chleba máslem.*
 Eng. a. Peter smeared butter on bread.
 b. Peter smeared bread with butter.

(2) a. *Turisté vylezli na kopec.*
 b. *Turisté vylezli kopec.*
 Eng. a. Tourists climbed up the hill.
 b. Tourists climbed the hill.

(3) a. *Petr řekl, že je Marie chytrá.*
 b. *Petr řekl o Marii, že je chytrá.*
 Eng. a. Peter said that Mary was clever.
 b. 'Peter - said - about - Mary - that - (she) is - clever.'

The uses of the verb *namazat* 'to smear', *vylézt* 'to climb' and *říci* 'to say' in (1a)-(1b), (2a)-(2b) and (3a)-(3b), respectively, refer to the same situations. However, they differ in their respective valency frames. We discuss three typologically different changes in verbal valency structure. We propose their adequate representation in the valency lexicon of Czech verbs, VALLEX.[1]

The VALLEX lexicon uses as its theoretical background the *Functional Generative Description* (henceforth FGD). In FGD, valency is related to the tectogrammatical layer, i.e., a layer

of linguistically structured meaning, (Sgall et al., 1986; Panevová, 1994). Valency characteristics of a verb are encoded in a form of valency frames. The frames are modeled as a sequence of valency slots, each valency slot standing for a single valency complementation. The slots consist of (i) a functor (rather coarse-grained tectogrammatical role labeling the relation of a complementation to a verb), (ii) information on obligatoriness, and (iii) a list of possible morhemic form(s), (Žabokrtský, Lopatková, 2007).

2 Situational vs. Structural Meaning

Let us turn back to examples (1), (2) and (3). The pairs of the uses *namazat* 'to smear' in (1a)-(1b), *vylézt* 'to climb' in (2a)-(2b) and *říci* 'to say' (3a)-(3b), respectively, denote the same situations, i.e., the individual situations portrayed by these uses are characterized by the same set of participants related by the same relations. We refer to this part of a verbal meaning as a *situational meaning* and to its components as *situational participants*, (Mel'čuk, 2004). The situational meaning represents such part of the verbal meaning which has not been syntactically structured yet. The part of the verbal meaning in which the components of the situational meaning is syntactically structured is referred here as a *structural meaning*. The structural meaning corresponds to the tectogrammatical layer and its components to the *valency complementations*, (Panevová, 1994).

Each lexical unit of a verb is characterized by both situational and structural meaning in a unique way: any change in the situational or structural part of its meaning leads to a change of lexical unit. We can observe that the pairs of the uses of the verbs *namazat* 'to smear' in (1a)-(1b), *vylézt* 'to climb' in (2a)-(2b) and *říci* 'to say' in (3a)-(3b), respectively, share the same situational meaning; however, they differ from each other in the structural part of meaning: The same set of sit-

[1]http://ufal.mff.cuni.cz/vallex/2.5/

uational participants are mapped onto the valency complementations in a different way, respectively, i.e., these uses of the verbs are characterized by different valency frames. It follows that the uses of the verbs *namazat* 'to smear' in (1a)-(1b), *vylézt* 'to climb' in (2a)-(2b) and *říci* 'to say' in (3a)-(3b), respectively, represent separate lexical units.

However, we observe that these separate lexical units are characterized by different types of the asymmetry in the correspondence between situational participants and valency complementations. We assume that the types of the asymmetry are closely related to characteristics of changes in a valency frame. As a consequence, these asymmetries determine a lexicographic representation of changes in a valency frame. For these reasons, the asymmetries represent a starting point in our analysis of changes in valency structure of verbs, in contrast to other approaches, e.g., (Levin, 1993).

Now let us focus on the verbs *namazat* 'to smear', *vylézt* 'to climb' and *říci* 'to say' again. On the basis of three types of the asymmetry, we determine three typologically different changes in the valency structure of these verbs: (i) semantic diatheses illustrated by the lexical units of the verb *namazat* 'to smear' in (1a)-(1b) (Section 3) (ii) multiple structural expression of a situational participant represented by the lexical units of the verb *vylézt* 'to climb' in (2a)-(2b) (Section 4) and (iii) structural splitting of a situational participant illustrated by the lexical units of the verb *říci* 'to say' in (3a)-(3b) (Section 5). We emphasize that all these changes are connected with separate lexical units, i.e., they are embedded in the lexical structure of the language.

In contrast, there are changes in valency frames which belong to the grammar structure of a language, (Apresjan, 1974). These changes are characterized by differences in the mapping of valency complementations onto surface syntactic positions while the correspondence of situational participants and valency complementations is preserved, (Kettnerová, Lopatková, 2010). With respect to the same situational and structural meaning, these changes are connected with different uses of a single lexical unit of a verb. As a consequence, different types of changes – the changes embedded in the grammar and lexical structure of a language – can be combined. For example, passive grammatical diathesis can be combined with the constructions in the relation of locative semantic diathesis,

as in (4a)-(4b):

(4) a. *Máslo bylo namazáno na chléb (od Petra).*
     b. *Chléb byl namazán máslem (od Petra).*
     Eng. a. Butter was smeared on bread (by Peter).
     b. Bread was smeared with butter (by Peter).

In this paper, we focus primarily on the changes in valency structure of verbs representing a part of the lexical structure of the language. The changes arising from the grammar structure of the language are left aside here. As the discussed changes in valency structure are based on the asymmetries in the correspondence between situational participants and valency complementations, an adequate representation of situational as well as structural meaning is necessary for the purpose of their description. However, whereas the representation of the structural meaning of verbs has been elaborated in detail in FGD, an adequate description of the situational meaning has not been worked up within this framework so far. Hence, we propose to enhance FGD (i) with lexical-semantic representation of the situational part of verbal meaning based on the lexical-conceptual structures, and (ii) with an open set of labels of situational participants.

3 Semantic Diatheses

The first type of the asymmetry in the mapping of situational participants and valency complementations can be illustrated by the uses of the verb *namazat* 'to smear' in (1a)-(1b). The situation denoted by this verb consists of three situational participants: 'Agent', 'Cover' and 'Surface'. The participants 'Cover' and 'Surface' can be mapped onto the valency complementations in two ways: 'Cover' onto PAT(ient) (1a) or EFF(ect) (1b) and 'Surface' onto DIR(ection) (1a) or PAT(ient) (1b). Thus either the participant 'Cover' or the participant 'Surface' are structured as the valency complementation PAT, which occupies the prominent syntactic position of object, see Figure 1. This fact results in a slight semantic difference between the uses in (1a) and (1b). This difference is associated with a holistic interpretation of the participant which is expressed as PAT, (Anderson, 1971; Dowty, 1991).

The asymmetry in the mapping is connected with the change of lexical unit of the verb, i.e., it is
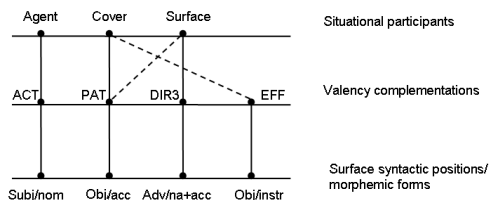
Figure 1: The asymmetry in the mapping of the situational participants and the valency complementations of the verb *namazat* 'to smear' associated with a semantic diathesis.

based on lexical-semantic means. We refer to the relations between such lexical units as semantic diatheses.[2] The changes in the valency structure of verbs associated with semantic diatheses may affect the number of valency complementations, their type and their morphemic form(s); moreover, these changes may vary even within one type of the diathesis. Thus we propose to represent separate lexical units related to a semantic diathesis by separate valency frames stored in the data component of the lexicon and to interlink them by a relevant type of the diathesis. E.g., two lexical units of the verb *namazat* 'to smear' are interlinked by the *locative semantic diathesis*.

In the grammar component, the changes in the mapping between situational participants and valency complementations are described by lexical rules based on an adequate lexical-semantic representation of the members of the semantic diathesis. For this purpose, we adopt the lexical-conceptual structure (henceforth LCS) proposed in (Rappaport Hovav, Levin, 1998). E.g., the uses of the verb *namazat* 'to smear' in (1a) and (1b) can be described by the following LCSs:

(a) $[[x \text{ ACT}<_{SMEAR}>] \text{ CAUSE } [\text{BECOME } [y \text{ ON } z]]]$

(b) $[x \text{ CAUSE } [\text{BECOME } [z <SMEARED>]] \text{ BY MEANS OF } [[x \text{ ACT}<_{SMEAR}>] \text{ CAUSE } [\text{BECOME } [y \text{ ON } z]]]]$

Commentary on the LCSs. LCS (a) corresponding to variant (1a) represents a change of location consisting of two subevents: (i) the first one represented as $[x \text{ ACT}<_{SMEAR}>]$ identifies the action of the 'Agent' x. The verb $<SMEAR>$ in the subscript serves as a modifier of the action. (ii) The

second part of the LCS $[\text{BECOME } [y \text{ ON } z]]$ represents the change of location of the 'Cover' y resulted from the first subevent, see the predicate CAUSE. In comparison with the LCS (a), the LCS (b) is more complex. In addition, it contains the component $[\text{BECOME } [z <SMEARED>]]$ specifying the change of state of the 'Surface' z indicated as $<SMEARED>$. Relating the component $[\text{BECOME } [z <SMEARED>]]$ with the whole LCS (a) indicates that this event arises as a consequence of the event identified by the LCS (a). The labels of the situational participants are associated with the position of the variables in the LCSs as follows: x ∼ 'Agent', y ∼ 'Cover', and z ∼ 'Surface'.

With respect to their complexity, we consider the LCS (a) as unmarked and the LCS (b) as marked. We formulate the following lexical rule Loc.r1 determining the change in the mapping between the situational participants and the valency complementations:

	LCS(a)		LCS(b)
y ∼ 'Cover'	PAT	$\Rightarrow_{Loc.r1}$	EFF
z ∼ 'Surface'	DIR	$\Rightarrow_{Loc.r1}$	PAT

Let us mention some other types of Czech semantic diatheses which can be represented in the lexicon in a similar way, i.e., by means of lexical rules determining the differences in the correspondence between situational participants and valency complementations:

Material-Product diathesis
    a. *Nařezal kládu*.PAT-Material *na tři polena*.EFF-Product.
    b. *Nařezal tři polena*.PAT-Product *z klády*.ORIG-Product
Eng. a. He cut the log.PAT-Material into three pieces.EFF-Product
b. He cut three pieces.PAT-Product from the log.ORIG-Material

Source-Substance diathesis
    a. *Slunce*.ACT-Source *vyzařuje teplo*.PAT-Substance
    b. *Teplo*.ACT-Substance *vyzařuje ze Slunce*.DIR-Source
Eng. a. The sun.ACT-Source radiates heat.PAT-Substance
b. Heat.ACT-Substance radiates from the sun.DIR-Source

Agent-Location diathesis
    a. *Včely*.ACT-Agent *se hemží na zahradě*.LOC-Location
    b. *Zahrada*.ACT-Location *se hemží včelami*.EFF-Agent

Eng. a. Bees.ACT-Agent are swarming in the garden.LOC-Location
b. The garden.ACT-Location is swarming with bees.EFF-Agent

4 Multiple Structural Expression of Situational Participant

The second type of the asymmetry in the correspondence between situational participants and valency complementations is illustrated by the uses of the verb *vylézt* 'to climb' in (2a) and (2b). The situation portrayed by this verb is composed by two situational participants: 'Agent' and 'Location'. In contrast to the semantic diatheses, this type of the asymmetry is not associated with any changes of situational participants in the prominent surface syntactic position of subject or object. Contrary, it results from two possible mappings of a single situational participant onto different valency complementations. In case of the verb *vylézt* 'to climb', the participant 'Location' is mapped either onto the valency complementation DIR(ection) (2a) or PAT(ient) (2b), see Figure 2.
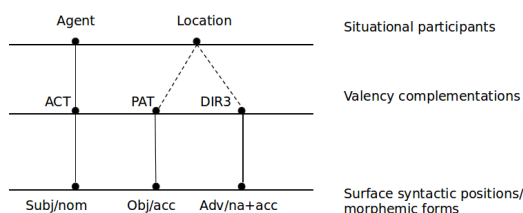

Figure 2: The multiple mapping of the situational participants 'Location' onto the valency complementations of the verb *vylézt* 'to climb'.

As in case of semantic diatheses, this type of the asymmetry, based on the lexical structure of the language, is connected with a change of lexical units of verbs. The changes in the valency structure of verbs can be described in a similar way as in case of semantic diatheses. E.g., in case of the verb *vylézt* 'to climb', separate lexical units characterized by different valency frames are stored in the data component of the lexicon. These frames are interlinked by a relevant type of the relation: In the grammar component, the lexical rule Dir.r1, based on the LCSs (c) and (d) (that correspond to (2a) and (2b), respectively), describes the change in the mapping of the situational participant 'Location'.

(c) [BECOME [x $<PLACE>$]]

(d) [BECOME [x $< PLACE_{TOP-OF}>$]]

Commentary on the LCSs. LCS (c) representing variant (2a) describes the change of location of the participant 'Agent' x. The location is identified with the constant $<PLACE>$. In contrast to variant (2b), an exact endpoint of the change of location is not implied here: 'Agent' can occur anywhere in the hill, including on the top of the hill. In the LCS (d) describing variant (2b), the constant is modified by the subscript specifying an endpoint of the change of location of 'Agent' – the top of the hill. With respect to the markedness of the constant in the LCS (d), we consider the LCS (c) as unmarked and the LCS (d) as marked. The label of the situational participant 'Agent' is attributed to the positions of the variables x in the LCSs. The situational participant 'Location' is implied by the constant $<PLACE>$.

The lexical rule Dir.r1 describing the change in the mapping of the situational participant 'Location' follows:

	LCS(c)		LCS(d)
'Location'	DIR	$\Rightarrow_{Dir.r1}$	PAT

Let us mention another type of multiple structural expression of a situational participant frequent in Czech, illustrated by the uses of the verb *umístit* 'to place'. In these uses of the verb, the situational participant 'Location' is mapped either onto the valency complementation DIR(irection) (5a) or onto the valency complementation LOC(ation) (5b). Two alternative mappings of this situational participant results in slightly different meanings: the event denoted by the first use of the verb *umístit* 'to place' in (5a) is conceived dynamically whereas the second use in (5b) is rather of static character:

(5) a. *Jana.*ACT-Agent *umístila dítě.*PAT-Patient *do jeslí.*DIR-Location
b. *Jana.*ACT-Agent *umístila dítě.*PAT-Patient *v jeslích.*LOC-Location
Eng. a. Jane.ACT-Agent placed her child.PAT-Patient into the nursery.DIR-Location
b. Jane.ACT-Agent placed her child.PAT-Patient in the nursery.LOC-Location

We assume that the above-mentioned example of multiple structural expression of the situational participant 'Location' may be described in the lexicon in a similar way, i.e., on the basis of a lexical rule determining two alternative mappings of the participant.

5 Structural Splitting of Situational Participant

The third type of the asymmetry in the correspondence between situational participants and valency complementations is illustrated by the verb *říci* 'to say' in (3a) and (3b). The situational participant 'Information' is mapped either onto one valency complementation EFF(ect) (3a) or onto two valency complementations PAT(ient) and EFF(ect) (3b), see Figure 3. We refer to these cases as a structural splitting of a situational participant.
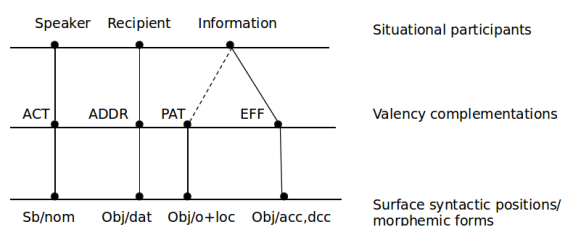

Figure 3: The structural splitting of the situational participant 'Information' of the verb *říci* 'to say'.

In case of this type of the asymmetry, the formulation of lexical rules describing changes in verbal valency structure is precluded as the coreferential relations between the split parts of a situational participant may be too complicated. See the following corpus example:

(6)   *A Šaron o Arafátovi řekl, že tomuto "králi vrahů a teroristů" nikdy nepodá ruku.*
Eng. 'Sharon - about - Arafat - said - that - this - "king of murderers and terrorists" - never - shakes - hand.'
(i.e., Sharon said [about Arafat] that he (=Sharon) would never shake hands with this "king of murderers and terrorist" (=Arafat's hands).)

For this reason, we propose to capture lexical units characterized by the structural splitting of a situational participant within a single valency frame in the data component of the lexicon. The split parts of a situational participant are represented by two co-indexed valency complementations. E.g., PAT and EFF corresponding to 'Information' are co-indexed by S in the valency frame of the verb *říci* 'to say':

$$\boxed{\text{ACT}_{obl} \ \text{ADDR}_{obl} \ \text{PAT}^S_{opt} \ \text{EFF}^S_{obl}}$$

This valency frame explicitly describes the use of the verb *říci* 'to say' with split 'Information',

as in (3b). In case of 'univalent' expression of 'Information' (3a), the situational participant corresponds to a more prominent valency complementation from the pairs of co-indexed valency complementations (e.g., in (3a) Information' is mapped onto EFF due to its obligatoriness).

In addition to the verbs of communication, the verbs denoting perception allow structural splitting of a situational participant:

(7)   a.   *Jana*.ACT-Perceiver *viděla, (jak Petr vchází do dveří)*.PAT-Phenomenon
b.   *Jana*.ACT-Perceiver *viděla Petra*.PAT-Phenomenon, *(jak vchází do dveří)*.EFF-Phenomenon
Eng. a.   'Jane.ACT-Perceiver - saw - (as - Peter - is entering - in - the doors)'.PAT-Phenomenon
b.   'Jane.ACT-Perceiver - saw - Peter.PAT-Phenomenon - (as - (he) is entering - in - the doors)'.EFF-Phenomenon
(i.e., Jane saw Peter entering the doors.)

Then the structural splitting of the situational participant 'Phenomenon' can be described in the lexicon in a similar way; i.e., both lexical units of the verb *vidět* 'to see' in (7a)-(7b) are represented by a single valency frame with co-indexed valency complementations corresponding to the split situational participant.

6 Conclusion

We distinguished three types of changes in valency structure of Czech verbs on the basis of three types of asymmetry in the correspondence between situational participants and valency complementations. We demonstrated that these changes, embedded in the lexical structure of the language, are connected with separate lexical units. In case of semantic diathesis and multiple structural expression of a situational participant, we proposed to represent these separate units by separate valency frames interlinked by a relevant type of the relation stored in the data component of the lexicon. Then in the grammar component, lexical rules determining the changes in the mapping between situational participants and valency complementations are formulated. In case of the structural splitting of a situational participant, possible complicated coreferential relations obstruct formulating lexical rules. Hence, we propose to represent both lexical units within a single valency frame with an abbreviated notation.

References

Stephen R. Anderson. 1971. On the Role of Deep Structure in Semantic Interpretation. *Foundations of Language*, 7:387–396.

Jurij D. Apresjan. 1974. *Leksicheskaja semantika. Sinonimicheskie sredstva jazyka.* Nauka, Moskva.

David Dowty. 1991. Thematic Proto-Roles and Argument Selection. *Language*, 67(3):547–619

Václava Kettnerová and Markéta Lopatková. 2010. The Representation of Diatheses in the Valency Lexicon of Czech Verbs. In H. Loftsson, E. Rögnvaldsson, S. Helgadóttir, eds.: *Proceedings of IceTAL 2010.* LNAI 6233, Berlin Heidelberg, Springer Verlag. 185–196.

Beth C. Levin. 1993. *English Verb Classes and Alternations: A Preliminary Investigation.* University of Chicago Press, Chicago, IL.

Igor A. Mel'čuk. 2004. Actants in Syntax and Semantics. *Language*, 12:1–66.

Jarmila Panevová. 1994. Valency Frames and the Meaning of the Sentence. In P. A. Luelsdorff, ed.: *The Prague School of Structural and Functional Linguistics.* John Benjamins Publishing Company, Amsterdam, Philadelphia. 223–243.

Malka Rappaport Hovav and Beth Levin. 1998. Building Verb Meanings. In M. Butt and W. Geuder, eds.: *The Projection of Arguments. Lexical and Compositional Factors.* SLI Publications, Stanford, California. 97–134.

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

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