The Rule-Based Approach
to Czech Grammaticalized Alternations⋆

Václava Kettnerová, Markéta Lopatková, and Zdeňka Urešová

Charles University in Prague
Institute of Formal and Applied Linguistics
{kettnerova,lopatkova,uresova}@ufal.mff.cuni.cz

Abstract. Under the term grammaticalized alternations, we understand changes in valency frames of verbs corresponding to different surface syntactic structures of the same lexical unit of a verb. Czech grammaticalized alternations are expressed either (i) by morphological means (diatheses), or (ii) by syntactic means (reciprocity). These changes are limited to changes in morphemic form(s) of valency complementations; moreover, they are regular enough to be captured by formal syntactic rules.

In this paper a representation of Czech grammaticalized alternations and their possible combination is proposed for the valency lexicon of Czech verbs, VALLEX.

1 Introduction

Prototypically, a single meaning of a verb can be surface syntactically structured in different ways. It follows that changes in valency frame of a verb (usually called alternations, see [1]) must be described either by syntactic rules, or they must be specified in lexicon entries. Here we focus on the changes resulting from the use of specific grammatical means (e.g., passivisation or reciprocity). These changes are referred here to as grammaticalized alternations. Czech grammaticalized alternations are expressed either (i) by morphological means (1a)–(1b) (traditionally referred to as diatheses), or (ii) by syntactic means (2a)–(2b) (reciprocity).

(1) a. *Mobilní operátoři snížili$_{active}$ cenu volání.*
   The mobile network operators reduced$_{active}$ the price of calls.
   b. *Cena volání se snížila$_{deagentive}$.*
   'The price of calls – *refl* – reduced$_{deagentive}$.'

(2) a. *Petr líbá Marii.*
   Peter kisses Mary.
   b. *Petr a Marie se líbají.*
   Peter and Mary kiss (each other).

Whereas the same type of grammatical means cannot be combined together, the combinations of diatheses and reciprocity are allowed within a single surface syntactic structure, see example of the verb *domluvit* 'to arrange' in (3a)–(3b):

(3) a. *Petr si domluvil$_{active}$ s Janem schůzku.*
    Peter arranged$_{active}$ an appointment with John.
  b. *Petr a Jan (spolu) mají domluvenu$_{resultative}$ schůzku.*
    Peter and John have arranged$_{resultative}$ an appointment.

The objective of this paper is to propose a representation of Czech grammaticalized alternations, i.e., diatheses and reciprocity, and their combinations in the valency lexicon of Czech verbs, VALLEX[1]. This lexicon takes the Functional Generative Description (FGD) as its theoretical background. In FGD, valency is related to the layer of linguistically structured meaning, see [2] and [3]. In VALLEX, valency characteristics of a verb are encoded in valency frames which are modeled as sequences of valency slots, each slot standing for a single valency complementation. The slots consist of a functor (coarse-grained semantic role), a list of morphemic form(s) and information on obligatoriness.

In order to satisfy needs of both human users and automated language processing, this lexicon is available in three formats: XML, HTML and PDF version. The information on valency of verbs (including the linguistically adequate and economic description of grammaticalized alternations) can be used for various NLP tasks, as e.g., machine translation, tagging, word sense disambiguation.

We demonstrate that grammaticalized alternations of both types (diatheses and reciprocity) are limited to changes in morphemic form(s) of valency complementations and that these changes are regular enough to be captured by formal syntactic rules. For the representation of combination of diatheses and reciprocity, additional rules are not required. However, explicit rule ordering is necessary to be determined.

From a broader perspective, we address a general question which part of the information needed for a language description is to be captured by general rules (i.e., in the grammar component of the language system) and which part is best recorded in the form of lexicon entries.

2 Grammaticalized Alternations: Theory

2.1 Diatheses

A central type of grammaticalized alternations is characteristic of the relations between surface syntactic structures which differ in the morphological category of voice. In Czech, these relations are referred to as diatheses. The members of diatheses are characterized by a permutation of valency complementations. This permutation affects the prominent surface position of subject from which the valency complementation 'ACTor' is prototypically shifted.

Five types of these relations are determined according to five marked morphological meanings of a verb: *passive*, *deagentive*, *dispositional*, *resultative*, and *recipient passive* diathesis. The syntactic structures characterized by these marked morphological meanings represent the marked members of diatheses whereas the structures with active voice of a verb constitute the unmarked one, see esp. [4].

---

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

2.2   Reciprocity

In contrast to diatheses, reciprocity represents rather peripheral type of grammaticalized alternations. It is connected with 'symmetricalization' – a syntactic operation conducted on two (or three) valency complementations, which (if their semantic properties allow for it) are used symmetrically; e.g., (2b) means that *Petr líbá Marii a (zároveň) Marie líbá Petra* 'Peter kisses Mary and (at the same moment) Mary kisses Peter', see esp. [5].

The reciprocal constructions are associated with the shift of the valency complementation expressed in a less prominent surface syntactic position into the more significant syntactic position (subject or direct object) of the other symmetrically used valency complementation. Whereas the prominent position is 'multiplied' (either by coordination, or plural), the less significant position is either deleted from the resulted surface syntactic structure, or it is realized with reflexive pronoun. Reflexive verb forms must be used if 'ACTor' is involved in the reciprocity relation.

Different types of reciprocity can be determined according to valency complementations which enter into the symmetric relation.

The reciprocal constructions can be considered as marked members of this type of alternation whereas the unreciprocal constructions as unmarked ones.

2.3   Combination of Diatheses and Reciprocity

Whereas the same type of grammaticalized alternations cannot be combined together, the combinations of diatheses and reciprocity are allowed within a single surface syntactic structure. Such combinations are applicable on condition that a certain morphological meaning (Section 2.1) can be applied to a verb and at the same time some of its valency complementations can be used symmetrically (Section 2.2). See the following example of the verb *domluvit* 'to arrange' allowing both reciprocal use of its valency complementations 'ACTor' and 'ADDRessee' (4b) and resultative morphological meaning (4c). Then these linguistic means can be combined within a single surface structure (4d):

(4) a. *Petr$_{ACT}$ domluvil$_{active}$ s Janem$_{ADDR}$ schůzku.*
  Peter arranged$_{active}$ an appointment with John.
  b. *(Petr a Jan)$_{rcp:ACT-ADDR}$ si (spolu) domluvili schůzku.*
  (Peter and John)$_{rcp:ACT-ADDR}$ arranged an appointment (together).
  c. *Petr má s Janem domluvenu$_{resultative}$ schůzku.*
  'Peter – has – with John – arranged$_{resultative}$ – an appointment.'
  d. *(Petr a Jan)$_{rcp:ACT-ADDR}$ (spolu) mají domluvenu$_{resultative}$ schůzku.*
  '(Peter and John)$_{rcp:ACT-ADDR}$ – (together) – have – arranged$_{resultative}$ –
  – an appointment.'

However, in some cases, the combination of a certain type of diathesis and reciprocity leads to an ungrammatical structure despite being separately available for a verb, see e.g. the verb *vyhubovat* 'to tell off'. Although this verb allows both for reciprocal use of 'ACTor' and 'PATient' (5b) and for recipient passive morphological meaning (5c), the combination of these linguistic means within a single surface syntactic structure results in an ungrammatical construction (5d).

(5) a. *Jan$_{ACT:Subj}$ vyhuboval$_{active}$ Marii$_{PAT:InObj}$.*

    John$_{ACT:Subj}$ told$_{active}$ Mary$_{PAT:InObj}$ off.

  b. *(Jan a Marie)$_{rcp:ACT-PAT:Subj}$ si (vzájemně) vyhubovali.*

    (John and Mary)$_{rcp:ACT-PAT:Subj}$ told off (each other).

  c. *Marie$_{PAT:Subj}$ dostala (od Jana)$_{ACT:Adv}$ vyhubováno$_{recipient}$.*

    'Mary$_{PAT:Subj}$ – got – (from John)$_{ACT:Adv}$ – told$_{recipient}$ off.'

  d. **dostali (vzájemně) vyhubováno$_{recipient}$ (od Jana a Marie)$_{rcp:ACT-PAT:Adv}$*

    'got – (each other) – told$_{recipient}$ off (by John and Mary)$_{rcp:ACT-PAT:Adv}$'

The reason of the ungrammaticality of the combination of recipient passive diathesis and reciprocity in (5d) lies in the fact that the surface syntactic shifts associated with these alternations are contradictory; formally, they result in the surface syntactic structure without subject while the verb form being inadequate. First, reciprocity (putting 'ACTor' and 'PATient' in symmetric relation) leads to the shift of 'PATient' (*Marii* in (5a)) to the subject position (which is multiplied by coordination in (5b)). Second, recipient passive diathesis prototypically consists in the following changes: (i) shifting the valency complementation occupying the subject to an adverbial position while (ii) the vacated subject being filled with the valency complementation corresponding to the cognitive 'Recipient' (expressed originally in dative); (iii) the dative surface position is deleted. However, by applying recipient passive diathesis on reciprocal construction in (5b), both (coordinated) 'ACTor' and 'PATient' (*Jan a Marie*) should be shifted from the subject which remains vacant as no dative 'Recipient' is present; as a result, the verb form is inappropriate for subject-less construction, see (5d).

Let us generalize this observation. Considering the basic postulates – (i) diatheses are characterized by 'ACTor' shifted from the subject syntactic position into a less prominent surface position (Section 2.1) whereas (ii) reciprocity is connected with the shift of a valency complementation occupying the less prominent surface syntactic position into the position of subject or direct object (Section 2.2), see [6], – we can formulate the following hypotheses:

A. If different types of grammaticalized alternations are combined, the order of their application can be prescribed; namely, in certain cases reciprocity should precede diathesis as diathesis may result in a surface construction that do not allow for certain types of reciprocity.

B. Moreover, the combinations of (a certain type of a) diathesis and reciprocity are allowed within a single surface syntactic structure under the condition that the application of reciprocity preserves formal conditions on the application of the particular diathesis.

3   Representation of Grammaticalized Alternations

For the purpose of the representation of grammaticalized alternations, we divide the lexicon into a data component and a rule component. In case of diatheses and reciprocity, the changes in the valency structure of verbs are limited to changes in morphemic form(s) of the valency complementations affected by the shifts in surface

syntactic positions; these changes are regular enough to be captured by formal syntactic rules. In Section 3.1, we demonstrate two examples of rules representing diathesis and reciprocity, respectively. For the description of these phenomena, we adopt syntactic rules formulated in the lexicon PDT-VALLEX, see esp. [7].[2] Section 3.2 is focused on the representation of the combination of diatheses and reciprocity.

3.1   Representation of Diatheses and Reciprocity

The **data component** of the lexicon contains lexical entries for individual lexical units. Only the unmarked valency frame, i.e., the valency frame representing use in the active unreciprocal structure, is stored for each lexical unit. A lexical unit is ascribed with the special attributes `-diat` and `-rcp` where the applicability of individual diatheses and reciprocity is specified, respectively. Then the **rule component** of the lexicon stores rules describing changes in valency frames associated with individual diatheses and reciprocity; in VALLEX, we make use of the rules designed for PDT-VALLEX [7].

Let us firstly demonstrate our approach on the example of passive diathesis of the verb *seznámit* 'to introduce'. The lexical entry of this verb stored in the data component is structured as follows (the lexical entry is simplified and translated for better understanding):

- **lemma:** **seznámit**^{pf} 'to introduce'
- **gloss:** *představit* 'to bring together'
- **frame:** ACT_1^{obl} ADDR_4^{obl} PAT_{s+7}^{obl}
- **example:** *seznámit přítele s příbuzným* 'to introduce the friend to the relative'
- **diat:** `pass, disp, res1, res2`
- **rcp:** ADDR-PAT, ACT-ADDR-PAT

In the rule component, the rule `Pass.r` given in Table 1 represents changes in valency structure of the verb associated with passive diathesis. (The rule is simplified here for better understanding; the thorough formal rule is split in order to cover all variants of relevant valency frames in the lexicon.)

Table 1. `Pass.r` rule for passive diathesis

Type	`passive`	
Action	verbform	`replace(active vf → passive vf)`
	ACT	`replace(nom → instr,od+gen)`
	ADDR	`replace(acc → nom)`

The `Pass.r` rule allows to derive the marked valency frame (6b) representing the passive surface structure (illustrated by example (7b)) from the valency frame (6a) corresponding to the unmarked active use of the verb (example (7a)).

---

[2] The rules (rule instances) in the cited work are a generalization of rules used in quality checking of the Prague Dependency Treebank 2.0 (PDT).

(6) a. verbform$_{active}$ ACT$^{obl}_{nom}$ ADDR$^{obl}_{acc}$ PAT$^{obl}_{s+instr}$ $\Rightarrow$

  b. verbform$_{passive}$ ACT$^{obl}_{instr,od+gen}$ ADDR$^{obl}_{nom}$ PAT$^{obl}_{s+instr}$

(7) a. *Sára*$_{ACT:Subj}$ *seznámila*$_{active}$ *přítele*$_{ADDR:Obj}$ *se svou matkou*$_{PAT:InObj}$.
  Sara$_{ACT:Subj}$ introduced$_{active}$ her friend$_{ADDR:Obj}$ to her mother$_{PAT:InObj}$.
  b. *Přítel*$_{ADDR:Subj}$ *byl* (*Sárou*$_{ACT:Adv}$) *seznámen*$_{passive}$ *s její matkou*$_{PAT:InObj}$.
  Her friend$_{ADDR:Subj}$ was introduced$_{passive}$ to her mother$_{PAT:InObj}$
   (by Sara$_{ACT:Adv}$).

The same principles can be applied on reciprocities. Only valency frames corresponding to unreciprocal structures are contained in the **data component**. By each relevant valency frame, valency complementations allowing for symmetrical use are listed in the special attribute -rcp. For example, the verb *seznámit* 'to introduce' allows for symmetrical use of 'ADDRessee' and 'PATient' (7c) or even of all three complementations 'ACTor', 'ADDRessee' and 'PATient' (7d). For the sake of simplicity, we limit our explanation to cases when only two valency complementations are affected.

(7) c. *Sára*$_{ACT:Subj}$ *seznámila* (*přítele a svou matku*)$_{rcp:ADDR-PAT:Obj}$.
  Sara$_{ACT:Subj}$ brought together (her friend and her mother)$_{rcp:ADDR-PAT:Obj}$.
  d. (*Sára, její přítel a matka*)$_{rcp:ACT-ADDR-PAT:Subj}$ *se* (*navázjem*) *seznámili*.
  (Sara, her friend and her mother)$_{rcp:ACT-ADDR-PAT:Subj}$ brought
   together (each other).

The **rule component** of the lexicon stores rules describing changes in valency frames associated with this type of reciprocity, see Table 2 (again, the rule is simplified here).

Table 2. `Rcp.r.ADDR-PAT` rule for reciprocity of 'ADDRessee' and 'PATient'

Type	rcp-ADDR-PAT		comment
Action	ADDR	`replace(acc → h-acc)`	(1)
	PAT	`replace(s+instr → !)`	(2)
		`add(`*spolu, navzájem, mezi sebou, …;* `opt)`	(3)

Commentary on the `Rcp.r.ADDR-PAT` rule:

(1) 'ADDRessee' stays in accusative and it must be realized by coordinated nouns, by plural noun or by an expression of semantic plurality (denoted by the symbol 'h').

(2) 'PATient' is merged with 'ADDRessee' and thus is not expressed in a separate surface syntactic position (denoted by '!').

(3) In addition, reciprocity may be lexically signalled by expressions such as *spolu* 'together', *navzájem* 'mutually', *mezi sebou* 'each other', etc.

The valency frame (8b) representing the reciprocal structure in (7c) can be derived from the valency frame (8a) corresponding to the unreciprocal structure (7a) on the basis of the rule given in Table 2:

(8) a. ACT$^{obl}_{nom}$ ADDR$^{obl}_{acc}$ PAT$^{obl}_{s+instr}$ $\Rightarrow$ b. ACT$^{obl}_{nom}$ ADDR-PAT$^{obl}_{h-acc}$

3.2   Representation of Combination of Diatheses and Reciprocity

We can observe that no additional rules are necessary for the representation of combinations of diatheses and reciprocity. However, in many cases explicit ordering of the rules describing the changes in valency frames must be determined. Esp. in cases where reciprocity involves 'ACTor' the rule representing this type of alternation must precede the rule(s) describing certain types of diatheses.

For instance, the passive diathesis and reciprocity of 'ADDRessee' and 'PATient' of the verb *seznámit* 'to bring together' can be combined within a single surface syntactic structure, see (7e):

(7) e. *(Její přítel a matka)*$_{rcp:ADDR-PAT:Subj}$ *byly (navzájem) představeni*$_{passive}$
(Sárou)$_{ACT:Adv}$.
(Her friend and her mother)$_{rcp:ADDR-PAT:Subj}$ were brought together$_{passive}$
(by Sara)$_{ACT:Adv}$.

The valency frame in (9c) describing combination of reciprocity 'PATient' and 'AD-DResee' and passive diathesis (as illustrated in (7e)) can be derived from the valency frame corresponding to active and unreciprocal structure (9a) by the consecutive application of the rule Rcp.r.ADDR-PAT (Table 2) and of a rule similar to the Pass.r (Table 1):

(9) a. verbform$_{active}$ ACT$^{obl}_{nom}$ ADDR$^{obl}_{acc}$ PAT$^{obl}_{s+instr}$ $\Rightarrow$

b. verbform$_{active}$ ACT$^{obl}_{nom}$ ADDR-PAT$^{obl}_{h-acc}$ $\Rightarrow$

c. verbform$_{passive}$ ACT$^{obl}_{instr,od+gen}$ ADDR-PAT$^{obl}_{h-nom}$

The explicit ordering of the rules representing diatheses and reciprocities is enforced by strict conditions imposed on their applications. In case that a lexical unit of a verb allows for applying a certain type of diathesis and at the same time some of its valency complementations can be used reciprocally, the respective rules for this type of diathesis and reciprocity are applied consecutively only in case that they meet the conditions of their applications; in other case this operation is blocked (see also Section 2.3).

4   Conclusion and Future Work

We have proposed the representation of Czech grammaticalized alternations associated with diatheses and reciprocities, with the focus on their combinations. We have demonstrated that these alternations can be described by formal syntactic rules. These rules are stored in the rule component of the lexicon. In the data component, only valency frames corresponding to the unmarked use (i.e., active unreciprocal use) of lexical units are captured; marked (morpho)syntactic uses of a single lexical unit are obtained by applying syntactic rules from the rule component. In cases when diatheses and reciprocity are combined, explicit rule ordering is necessary to be determined.

The proposed representation of Czech grammaticalized alternations has been already partially applied in the valency lexicon of Czech verbs VALLEX (namely deagentive diathesis and reciprocity). Further enrichment of the lexicon is planned for future in two directions: (i) Full adaptation of syntactic rules defined for PDT-VALLEX is under

preparation.[3] (ii) Inspired by [8], (semi)automatic identification of the lexical items in the data component that allow for individual types of diatheses is in development. From the theoretical point of view, reciprocity affecting three valency complementations deserves further attention. Moreover, special attention will be paid to the possible combinations of the grammaticalized alternations.

Only the implementation of the tasks mentioned above – the adaptation of formal syntactic rules and identification of relevant lexical items for individual diatheses – allows for thorough evaluation of the outputs of the proposed rules (in a form of valency frames corresponding to marked (morho)syntacic uses of lexical units of Czech verbs). This will be of a great interest since the original rules designed for PDT consistency checking 'over-generated' (in a sense they allowed also wrong surface configurations, relying on the fact that the underlying text analyzed in the corpus was grammatically correct Czech). We suppose that this shortcoming will be eliminated by imposing strict conditions on the application of the rules.

References

1. Levin, B.C.: English Verb Classes and Alternations: A Preliminary Investigation. The University of Chicago Press, Chicago (1993)
2. Sgall, P., Hajičová, E., Panevová, J.: The Meaning of the Sentence in Its Semantic and Pragmatic Aspects. Reidel, Dordrecht (1986)
3. 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 Publishing Company, Amsterdam (1994)
4. Panevová, J., et al.: Syntax současné češtiny (na základě anotovaného korpusu). Nakladatelství Karolinum, Praha (manuscript)
5. Panevová, J., Mikulová, M.: On Reciprocity. The Prague Bulletin of Mathematical Linguistics 87, 27–40 (2007)
6. Kettnerová, V., Lopatková, M.: Changes in valency structure of verbs: Grammar vs. lexicon. In: Levická, J., Garabík, R. (eds.) Proceedings of Slovko 2009, NLP, Corpus Linguistics, Corpus Based Grammar Research, Bratislava, SAV, pp. 198–210 (2009)
7. Urešová, Z.: Valence sloves v Pražském závislostním korpusu. In: Studies in Computational and Theoretical Linguistics, Institute of Formal and Applied Linguistics, Prague (2011)
8. Skoumalová, H.: Czech Syntactic Lexicon. Ph.D. thesis, Charles University in Prague (2001)

---

[3] The total number of rules described in [7] is 44; since some of the rules serve as templates over varying functors, the number of rule instances is over 100.