

Reflexive Verbs in a Valency Lexicon: The Case of Czech Reflexive Morphemes

Václava Kettnerová, Markéta Lopatková

Charles University in Prague

E-mail: kettnerova@ufal.mff.cuni.cz, lopatkova@ufal.mff.cuni.cz

Abstract

In this paper, we deal with Czech reflexive verbs from the lexicographic point of view. We show that the Czech reflexive morphemes *se* and *si* constitute different linguistic meanings: either they are formal means of the word formation process of the so called reflexivization, or they are associated with the syntactic phenomena of reflexivity, reciprocity, and diatheses.

All of these processes are associated with changes in the valency structure of verbs. We formulate a proposal for their lexicographic representation for the valency lexicon of Czech verbs, VALLEX. We make use of the division of the lexicon into a data component and a grammar component which represents a part of the overall Czech grammar. The data component stores information on valency structure of verbs in unmarked (active) structures. The grammar component consists of formal rules describing regular changes in the valency structure of verbs; these rules allow for the derivation of valency frames underlying the usages of verbs in marked structures (reflexive, reciprocal, deagentive and dispositional) from the valency frames corresponding to unmarked structures (non-reflexive, unreciprocal, and active).

Czech reflexive verbs thus represent an illustrative example of the lexical-grammar interplay: we demonstrate that a close interaction between the lexicon and the grammar is necessary for a representation of these verbs and they both are indispensable if such a representation is to be adequate and economical.

Keywords: reflexive verb, reflexive morpheme, valency lexicon, Czech

1 Introduction

In this paper, the possibility of a lexicographic representation of the reflexivity of Czech verbs is described in detail. In Czech, the reflexives *se* and *si* are on the one hand formal means of word formation process of so called reflexivization; on the other hand, they are associated with syntactic phenomena of reflexivity (in the narrow sense, also called “true reflexives”), reciprocity, and diatheses. According to their function, the reflexives *se* and *si* represent clitic morphemes corresponding either (ia) to components of verb lemmas (*se* and *si*), or (ib) to a component of verb form (only *se*), or (ii) to the personal pronoun *se* (with its inflected variant *si* and corresponding non-clitic variants *sebe* and *sobě*, respectively). As examples (1), (2), (3) and (4) with the verb *zabít* “to kill” show, both types of reflexives can occur with a single verb, constituting different linguistic meanings: in example (1), *se* is a component of the verb lemma *zabít se* “to kill (oneself)” (type (ia)); in examples (2) and (3), *se* is interpreted as the reflexive personal pronoun (however, expressing different meanings, true reflexivity and reciprocity, respectively, type (ii)); and example (4) illustrates *se* as a component of a verb form of the verb *zabít* “to kill” (type (ib)).

- (1) *Zabil se pádem ze střechy.* (CNC, SYN2006pub)
“killed – SE_{morph} – by falling – from roof.”
Eng. He killed himself (unintentionally) by falling from the roof.
- (2a) *Zabil se vlastní zbraní ...* (CNC, SYN2006pub)
“killed – SE_{pron} – own – weapon ...”
Eng. He killed himself with his own weapon ...
- (2b) *Zabil sebe vlastní zbraní.*
“killed – SEBE_{pron} – own – weapon ...”
Eng. He killed himself with his own weapon ...
- (3a) *Zabili se navzájem.*
“killed – SE_{pron} – each other.”
Eng. They killed each other.
- (3b) *Zabili sebe navzájem.*
“killed – SEBE_{pron} – each other.”
Eng. They killed each other.
- (4) *... zabila se dvě vykrmená prasata a pečínka provoněla celý dům.* (CNC, SYN2005)
“... killed – SE_{morph} – two – fattened – pigs – and – roast meat – scented – the whole – house.”
Eng. Two fattened pigs were killed and roast meat scented the whole house.

We show that the reflexives *se* and *si* belong to several different language phenomena which involve specific changes in the valency structure of verbs. As a consequence, they require to be represented in a lexicon in different ways. Here we describe the representation of these phenomena in the Valency Lexicon of Czech Verbs, VALLEX.

1.1 Related Work

Reflexivity has been extensively studied in the theoretical linguistics since the 1980s. The research has focused on linguistic means encoding reflexivity, their interpretations and ambiguities in individual languages. Recently, this linguistic phenomenon has received considerable attention even from the cross-linguistic perspective (König & Gast, 2008), (Nedjalkov, 2007). Numerous analyses show that linguistic means expressing reflexivity are usually ambiguous as they fulfill diverse functions in individual languages and that drawing clear distinctions between these functions represents a tricky task. For these reasons, developing a satisfactory lexicographic representation of reflexivity – despite being highly beneficial esp. for natural language processing and foreign learners – remains rather challenging (Renau & Battaner, 2012).

In Czech, reflexivity encoded by the reflexives *se* and *si* represent widely debated phenomenon from both theoretical (Oliva, 2001; Panevová, 1999, 2007) and computational point of view (Oliva, 2003; Petkevič, 2013). Two primary functions of the Czech reflexives are determined: (i) the reflexives as components of verb lemmas or verb forms and (ii) the reflexives as the personal pronoun, see Section 1. However, despite the plenitude of studies focused on Czech reflexivity, testable criteria for their distinction have not yet been established. In most cases, the substitutability of the reflexives *se* and *si* with *sebe* and *sobě*, respectively, can be applied as an operational test for distinguishing the reflexive personal pronoun from *se* and *si* as the components of verb lemmas or verb forms. However, this test can fail esp. in cases where the substitution leads to stylistically unacceptable sentences or in cases of haplology, see Section 2.1. In such cases, we adopt solutions taking economy and systematicity of the lexicographic representation into account.

As we attempt a lexicographic representation of reflexivity in a lexicon, let us introduce several lexical resources providing the information on these phenomena. First, *LexIt*, a large-scale lexical resource providing the automatically derived information on subcategorization and semantic properties of Italian verbs, nouns and adjectives stores the information on reflexivity as well (Lenci et al., 2012). Second, this type of information is also covered in *Diccionario de enseñanza del español como lengua extranjera, DAELE*, a Spanish learner's dictionary (Renau & Battaner, 2012). Third, *FrameNet* records the information on reciprocity of frame elements (linguistic phenomenon closely related to reflexivity) by adding special semantic frames indicating reciprocity (Ruppenhofer, et al., 2010).

For Czech, *PDT-VALLEX*, a valency lexicon linked with word occurrences in the Prague Dependency Treebank 2.0 (PDT) (Hajič, et al., 2006), provides the information on valency behavior of verbs, nouns, adjectives and adverbs (Hajič, et al., 2003). Although the information on reflexivity and reciprocity of verbs is not explicitly recorded in this lexicon, it can be easily extracted from PDT (if reflexive or reciprocal usages appear in the corpus). In addition, a fully automatically derived *Czech Syntactic Lexicon* (which is however not publically available) was designed, providing the information on possible reciprocity, reflexivity and diatheses of verbs (Skoumalová, 2001).

1.2 VALLEX

The valency lexicon of Czech verbs, *VALLEX*,¹ is a collection of linguistically annotated data and documentation (Žabokrtský & Lopatková, 2007; Lopatková et al., 2008). It provides the information on valency structure of Czech verbs in their particular meanings / senses, possible morphological forms of their valency complementations and additional syntactic information accompanied with glosses and examples. In *VALLEX*, version 2, there are roughly 2,730 lexeme entries containing together around 6,460 lexical units ('senses'). Verb lexemes were selected according to their frequency in the Czech National Corpus.² The lexicon has been developed for both human users and NLP applications, and is therefore in three different formats: HTML, XML and printable versions.

In *VALLEX*, the valency theory developed within the theoretical framework of the Functional Generative Description (henceforth FGD) is used as the theoretical background for the description of valency of verbs, see esp. (Sgall et al., 1986), (Panevová, 1994). According to this theory, valency complementations are divided into arguments (inner participants) and free modifications (adjuncts). They both can be obligatory or optional. The types of (verbal) arguments are distinguished mainly on the basis of the syntactic behavior of verbs. Five types of arguments have been determined – 'Actor' (ACTor, label ACT), 'Patient' (PATient, PAT), 'Addressee' (ADDRessee, ADDR), 'Origin' (ORIGIn, ORIG), and 'Effect' (EFFect, EFF). In contrast to the arguments, free modifications are semantically distinctive, being identified on the basis of their syntactico-semantic functions.

In *VALLEX*, the key information on the valency structure of a given lexical unit is encoded in the form of valency frames. A valency frame is formed as a sequence of slots; each slot stands for one valency complementation and consists of its type ('ACTor', 'ADDRessee', etc.), possible morphemic forms and its obligatoriness (obligatory or optional). Further, each lexical unit can be characterized by additional syntactic information on, e.g., syntactico-semantic class membership, diatheses, reciprocity of valency complementations, reflexivity. This information is provided in special attributes attached to individual lexical units.

The lexicon is divided into the data and the grammar component; the latter stores rules describing regular syntactic properties of verbs and it represents a part of the overall grammar of Czech, see esp. (Kettnerová et al., 2012a). The close interplay of these two parts of the lexicon is demonstrated on the representation of the reflexives *se* and *si* in the following sections. First, the reflexives *se* and *si* as components of verb lemmas (i.e., as formal means of the word formation process of reflexivization) are discussed in detail, esp. the syntactic properties of reflexivization are described and their lexicographic description is outlined in Section 2. Second, the reflexive pronoun *se* as a formal means of reflexivity (in the narrow sense) and reciprocity is surveyed and its representation in the lexicon is introduced in Section 3. Finally, the

¹ <http://ufal.mff.cuni.cz/vallex/>

² <http://www.korpus.cz/>

reflexive *se* as a component of reflexive verb forms that is involved in two types of Czech diatheses is debated in Section 4. In conclusion, the lexical entry of the verb *zabít* “to kill” (illustrated in Section 1) is displayed.

2 Czech Morphemes *se* and *si* as Components of Verb Lemmas

In Czech, the reflexive morphemes *se* and *si* can represent *freestanding components of verb lemmas* of the so called *reflexive verbs*. In Czech, there are two types of reflexive verbs: (i) reflexive tantum verbs (Section 2.1) and (ii) derived reflexive verbs (Section 2.2).

2.1 Reflexive Tantum Verbs

The first type is represented by the so-called reflexive tantum verbs, i.e., the verbs that have no non-reflexive counterparts, e.g., *bát se* “to be afraid”, *smát se* “to laugh”, *stěžovat si* “to complain”, *zapamatovat si* “to remember”, *domnívat se* “to assume”, *chlubit se* “to boast”, *líbit se* “to like”, *ptát se* “to ask”, *zamilovat se* “to fall in love”, see examples (5) and (6). In the case of reflexive tantum verbs, the reflexive morpheme *se* or *si* is a part of the verb lemma representing the respective verb lexeme in the data component of the lexicon.

Moreover, there are cases in Czech where a reflexive verb seemingly has a non-reflexive counterpart but these reflexive and non-reflexive verbs are not related by any derivational relation: the lexical meanings of these verbs are completely different, e.g. *dít se* “to happen” vs. *dít* “to tell”, *dopustit se* “to commit” vs. *dopustit* “to fill (with water)”, *hodit se* “to match” vs. *hodit* “to throw”, see examples (7)-(8). These are represented as separate verb lexemes (in separate lexical entries) in the valency lexicon.³

- (5a) *Jan se bojí zkoušky.*
“John – SE_{morph} – is afraid – of the exam.”
Eng. John is afraid of the exam.
- (5b) **Jan bojí zkoušky.*
“John – is afraid – of the exam.”
- (6a) *Hosté si stěžovali na špatnou stravu v hotelu.*
“guests – SI_{morph} – complained – of bad food – at the hotel”
Eng. The guests complained about bad food at the hotel.
- (6b) **Hosté stěžovali na špatnou stravu v hotelu.*
“guests – complained – of bad food – at the hotel”
- (7) *Co se děje?*
“what – SE_{morph} – happens.”
Eng. What is happening?
- (8) *“Pravdu říkáš,” odpověděl Petr.*
“the truth – you are telling – replied – Peter.”
Eng. “You are telling the truth,” replied Peter.

2.2 Derived Reflexive Verbs

Reflexive verbs of the second type are derived from non-reflexive verbs by adding the freestanding morpheme *se* or *si*. This process is called reflexivization, see esp. (Dokulil, 1986). In Czech, the reflexivization is a productive word formation process, which is largely syntactically motivated. Basically, two types of changes in valency structure of verbs are associated with this process (Sections 2.2.1 and 2.2.2). Further, in rare cases, reflexivization does not involve any change in valency structure (Section 2.2.3).

2.2.1 Reflexivization Applied to Transitive Verbs Resulting in Reflexive Intransitive Verbs

When reflexivization is applied to transitive verbs, it results in reflexive intransitive verbs associated with specific shifts in the lexical meaning of verbs: whereas non-reflexive transitive verbs express intentional acts (9a), reflexive intransitive verbs prototypically indicate non-intentional acts (9b):⁴ the argument corresponding to the direct object (expressed by the accusative) of the transitive non-reflexive verb maps onto the subject (expressed by nominative) of the derived intransitive reflexive verb.

- (9a) *Maminka vaří brambory.*
“mother – cooks – potatoes_{Dobj-acc.}.”
Eng. The mother is cooking potatoes.

³ In case where a sentence contains more than one reflexive tantum verbs, the reflexive *se* can be subject to haplology: a single occurrence of the reflexive can be associated with two verbs, see the following example where both the verb *pokusit se* “to try” and *usmát se* “to smile” are reflexive tantum verbs:

Jan se pokusil usmát.
“John – SE_{morph} – tried – smile.”
Eng. John tried to smile.

⁴ The act expressed by verbs denoting movement can be conceived as intentionally or unintentionally performed, e.g., *Petr opřel kolo o zed’.* Eng. Peter leaned the bike against the wall. (intentional act) – *Petr se opřel o zed’.* Eng. Peter leaned against the wall. (un/intentional act).

- (9b) *Brambory se vaří.*
 “potatoes_{Subj-nom} – SE_{morph} – cook.”
 Eng. Potatoes are cooking.

2.2.2 Reflexivization Applied to Verbs Implying Reciprocity

Reflexivization is also involved in the derivation of reflexive reciprocal verbs, i.e., verbs indicating reciprocity in their lexical meaning (Panevová & Mikulová, 2007). These verbs are derived by the reflexive morphemes *se* or *si* from verbs that imply (at least two) semantically homogeneous arguments, typically structured as ACTor (in nominative) and PATient or ADDRessee (expressed either by the accusative or by the dative), see examples (10a) and (11a), respectively. (As for reflexive pronouns in reciprocal constructions, see esp. Section 3.2.)

Reflexive verbs indicating reciprocity are associated with specific changes in their valency structure: the argument of the non-reflexive verb that is expressed in the accusative or dative is expressed by a prepositional group with the reflexive verb indicating reciprocity, see examples (10b) and (11b), respectively.

- (10a) *Petr potkal Marii.*
 “Peter – met – Mary_{PAT-acc.}”
 Eng. Peter met Mary.
- (10b) *Petr se potkal s Marií.*
 “Peter – SE_{morph} – met – with Mary_{PAT-s+instr.}”
 Eng. Peter met with Mary.
- (11a) *Dědeček vypráví dětem pohádky.*
 “the grandpa – tells – the children_{ADDR-dat} – fairy tales.”
 Eng. The grandpa is telling the children fairy tales.
- (11b) *Dědeček si vypráví s dětmi pohádky.*
 “the grandpa – SI_{morph} – tells – with the children_{ADDR-s+instr} – fairy tales”
 Eng. The grandpa and the children are telling each other fairy tales.

2.2.3 Reflexivization without Changes in Valency Structure

For a limited number of verbs, reflexivization does not result in any changes in either the valency structure or the meaning. The derivation by the morphemes *se* or *si* without clear syntactic or semantic motivation can be illustrated by the following examples (12) and (13).

- (12a) *Myslím, že je to dobře.*
 “I think – that – is – it – good.”
 Eng. I think that it is good.
- (12b) *Myslím si, že je to dobře.*
 “I think – SI_{morph} – that – is – it – good.”
 Eng. I think that it is good.
- (13a) *Zítřa začíná ve městě festival vína.*
 “tomorrow – starts – in the town – a festival of wine.”
 Eng. Tomorrow a wine festival starts in the town.
- (13b) *Zítřa se ve městě začíná festival vína.*
 “tomorrow – SE_{morph} – in the town – starts – a festival of wine.”
 Eng. Tomorrow a wine festival starts in the town.

2.3 Representation of Reflexive Tatum and Derived Reflexive Verbs in the Lexicon

In the case of both *reflexive tantum verbs* and *derived reflexive verbs*, the reflexive morphemes *se* and *si* are represented in the data component of the lexicon as a part of their verb lemmas (Section 2.1, 2.2.1 and 2.2.2, respectively). Derived reflexive verbs and their non-reflexive counterparts are recorded as separate verb lemmas (and thus separate lexical entries). Only derived reflexive verbs without syntactic changes (Section 2.2.3) are handled as variants of the respective non-reflexive verbs.

3 Czech Morphemes *se* and *si* as a Reflexive Pronoun

The reflexive *se* can also represent a *personal pronoun* (with the morphemic form *se* for accusative and *si* for dative, and their non-clitic variants *sebe* and *sobě*, respectively). The reflexive pronoun expresses reflexivity (in the narrow sense, Section 3.1) and reciprocity (Section 3.2).

3.1 Reflexivity

In cases where ACTor performs an action that is focused on himself/herself (also called “true reflexivity”), the reflexive pronoun *se* is used in Czech as a formal means of grammatical coreference, see esp. (Hajičová, et al., 1985, 1986, 1987): in these cases, the reflexive pronoun *se* stands for an argument of the verb that is referentially identical with ACTor in the subject position, examples (14) and (15). The form of the reflexive pronoun depends on the morphemic case of the

argument (*se* in the accusative and *si* in the dative). In the case of reflexivity, the clitic forms of the reflexive pronoun *se/si* can be replaced by their non-clitic variants *sebe/sobě*:⁵

(14a) *Petr se myje.*

“Peter_{ACT-Subj} – SE_{pron-acc} – washes.”

Eng. Peter is washing himself.

(14b) *Petr myje sebe (ale ne dítě).*

“Peter_{ACT-Subj} – washes – SEBE_{pron-acc} – (but not the child).”

Eng. Peter is washing himself (but not the child).

(15a) *Marie si koupila k obědu sendvič.*

“Marie_{ACT-Subj} – SI_{pron-dat} – bought – for lunch – a sandwich.”

Eng. Mary bought herself a sandwich for lunch.

(15b) *Sobě k obědu Marie koupila sendvič, dětem hranolky.*

“SOBĚ_{pron-dat} – for lunch – Marie_{ACT-Subj} – bought – a sandwich, – to the children – French fries”

Eng. Mary bought a sandwich to herself and French fries to children for lunch.

Reflexivity is represented in the lexicon by a special attribute *-rfl* attached to relevant lexical units. In this attribute, the information about the possibility of the reflexive usage of some arguments is provided by the value *cor3* (for arguments in the dative, example (15)) and *cor4* (for arguments in the accusative), example (14)). Other forms (e.g., prepositional groups) are not explicitly marked in the lexicon as they are expressed only by long variants of the reflexive personal pronoun (which are not ambiguous).

3.2 Reciprocity

Further, the reflexive pronoun *se* can express reciprocity. Reciprocalization is a syntactic operation on two (or three) arguments of a verb which puts the involved arguments in the symmetry. The main conditions imposed on such arguments are (i) their semantic homogeneity and (ii) same status with respect to topic-focus articulation. Reciprocalization leads to specific changes in the valency structure of a verb: the involved argument expressed in a less prominent surface syntactic position is shifted to the more significant position (subject or direct object) of the other symmetrically used argument, see (Panevová, 1999, 2007) and (Panevová & Mikulová, 2007). The resulting surface syntactic structure is characterized by a “multiplied” subject (or direct object) which is filled by a coordination, example (16), morphological, example (17), or semantic plural (e.g., the collective noun in example (18)). The syntactic position of the shifted (less significant) argument is typically formally filled by the reflexive pronoun *se* (expressed in the appropriate case), see below.

(16a) *Petr a Pavel se bijí.*

“Peter – and – Paul – SE_{pron-acc} – beat.” “

Eng. Peter and Paul are beating each other.

(16b) *Petr a Pavel bijí sebe navzájem.*

“Peter – and – Paul – beat – SEBE_{pron-acc} – each other.”

Eng. Peter and Paul are beating each other.

(17a) *Děti se bijí.*

“children – SE_{pron-acc} – beat.”

Eng. Children are beating each other.

(17b) *Děti bijí sebe navzájem.*

“children – beat – SEBE_{pron-acc} – each other.”

Eng. Children are beating each other.

(18a) *Celá rodina si pomáhá.*

“whole – family – SI_{pron-dat} – help

Eng. The whole family helps each other.

(18b) *Rodina pomáhá sobě (navzájem – a ne jim).*

“family – helps – SOBĚ_{pron-dat} (each other – and not them).”

Eng. The family helps each other.

In Czech, reciprocal constructions are created by two different types of verbs, by non-reciprocal verbs, i.e., by verbs that do not imply reciprocity in their lexical meaning (Section 3.2.1), and by inherently reciprocal verbs (Section 3.2.2).

3.2.1 Verbs Not Implying Reciprocity

Many verbs in Czech can potentially express reciprocity although reciprocity is not implied in their lexical meaning,⁶ e.g., *děkovat* “to thank”, *obviňovat* “to accuse”, *hrozit* “to threaten”, *pomáhat* “to help”, *vydírat* “to blackmail”, examples (19) and (20). In such cases, the reciprocal constructions (as described above) are optionally accompanied with the lexical expressions *vzájemně*, *navzájem*, *jeden druhý* “each other”, and *spolu* “together”, emphasizing the reciprocal meaning.

⁵ The use of clitic and non-clitic variants of the reflexive pronoun is affected esp. by the topic-focus articulation – thus the possibility to replace clitic forms by non-clitic forms of the reflexive pronoun in a sentence is often conditioned by changes in word order; however, this issue is not addressed in this paper as it goes beyond its scope.

⁶ Compare also with Section 2.2.2 describing derived reflexive verbs that imply reciprocity in their lexical meaning.

- (19) *Manželé se (vzájemně) obviňují z nevěry.*
 “husband and wife – SE_{pron-acc} – (each other) – accuse – of infidelity.”
 Eng. Husband and wife accuse each other of infidelity.
- (20) *Otec a syn si (vzájemně) lhali, aby si neublížili.*
 “father and son – SI_{pron-dat} – (each other) – lied – in order to – SI_{pron-dat} – not-to-hurt.”
 Eng. Father and son lied (to each other) in order not to hurt each other.

The lexical expressions (explicitly) indicating reciprocity are, however, obligatory in reciprocal constructions created by reflexive tantum verbs that do not imply reciprocity. For instance, although the verbs *smát se* and *vysmívat se* “to laugh at” do not imply reciprocity, their ACTor and ADDRessee can be put in the symmetrical relation. In reciprocal constructions with these verbs, *se* represents the morpheme that is a component of the verb lemmas, not the reflexive pronoun (as it cannot be substituted with the non-clitic form *sebe*). The reciprocity therefore must be expressed by lexical means, see example (21a) and (22a). If no such lexical means is present, the construction is either not reciprocal (21b), or even not grammatical (22b):

- (21a) *Petr a Pavel se smáli jeden druhému.*
 “Peter and Paul – SE_{morph} – laughed – at each other.”
 Eng. Peter and Paul were laughing at each other.
- (21b) *Petr a Pavel se smáli.*
 “Peter and Paul – SE_{morph} – laughed.”
 Eng. Peter and Paul were laughing.
- (22a) *Petr a Pavel se vysmívali jeden druhému.*
 “Peter and Paul – SE_{morph} – laughed – at each other.”
 Eng. Peter and Paul were laughing at each other.
- (22b) **Petr a Pavel se vysmívali.*
 “Peter and Paul – SE_{morph} – laughed.”

3.2.2 Verbs Implying Reciprocity

In addition, reciprocalization can be also applied to inherently reciprocal verbs. There are basically two types of such verbs: (A) *verbs indicating reciprocity in their lexical meaning* that might undergo the derivation of reflexive verbs indicating reciprocity (see Section 2.2.2) and (B) *non-reflexive verbs that imply reciprocity* of some of their arguments *in their lexical meaning* (e.g., *soupeřit* “to fight”, *sousedit* “to neighbor”).

In the case of (A), in addition to the derivation of reflexive reciprocal verbs (with the change of an accusative or dative complement into a prepositional group (see Section 2.2.2 and example (10), here repeated as (23)), the verbs can undergo “standard” reciprocal derivation, as in example (24a).

Here the question arises whether reciprocal constructions are derived from the non-reflexive *potkat* “to meet” (23a) or the reflexive verb *potkat se* “to meet” (23b). From the theoretical point of view, this question remains still open. The reflexive *se* in these constructions can be seen as the reflexive pronoun (as it can be replaced by its non-clitic variant *sebe* (24b)); however, the possible haplology of the morpheme *se* (Petkevič, 2013) makes the interpretation complicated and the theoretical interpretation of the derivation of these reciprocal constructions can differ, see also (Panevová, 2007).

For the practical implementation in the lexicon, we propose to derive the reciprocal constructions as in (24a) from the non-reflexive verbs (*potkat* for this case), not from its reflexive counterpart (*potkat se*) since this proposal allows us to use a single derivational rule for both types of verbs allowing for reciprocity (Section 3.2.1 and 3.2.2A).

- (23a) *Petr potkal Marii.*
 “Peter – met – Mary_{PAT-acc}.”
 Eng. Peter met Mary.
- (23b) *Petr se potkal s Marií.*
 “Peter – SE_{morph} – met – with Mary_{PAT-s+instr}.”
 Eng. Peter met with Mary.
- (24a) *Petr a Marie se potkali (navzájem).*
 “Peter and Mary – SE_{pron-acc} – met (each other)”
 Eng. Peter and Mary met (each other).
- (24b) *Petr a Marie potkali sebe navzájem i další přátele.*
 “Peter and Mary met – SEBE_{pron-acc} – each other – and other friends.”
 Eng. Peter and Mary met (each other) as well as other friends.

In the case of (B), *non-reflexive verbs*, changes in the valency structure (i.e., the “multiplication” of subject) are sufficient markers of reciprocity, and the reflexive pronoun is not present in the reciprocal structures (25).

- (25) *Týmy ČR soupeří o postup do finále. (= tým s týmem // mezi sebou/spolu)*
 Eng. The teams of the Czech Republic fight for the finals. (= each team with other teams)

3.3 Reciprocity in the Lexicon

Reciprocity of arguments is described in the data component of the lexicon in a special attribute *-rcp* providing the list of the arguments that can enter the symmetrical relation. Changes in the valency structure of verbs (including the use of lexical means for expressing reciprocity) are regular enough to be captured by formal rules. These rules are stored in the grammar component of the lexicon and they make it possible to automatically derive valency frames underlying reciprocal constructions, see (Kettnerová et al., 2012b).

4 Czech Morpheme *se* as a Component of a Reflexive Verb Form

Finally, the reflexive *se* can represent a freestanding *morpheme* that is a component of a reflexive verb form in two types of diatheses in Czech: (i) the deagentive diathesis (Section 4.1) and (ii) the dispositional diathesis (Section 4.2). Diatheses are relations between syntactic structures of a verb which differ in the grammatical category of voice, i.e., they are associated with specific morphological meanings of a verb.

In Czech, five specific morphological meanings are determined: passive, deagentive, resultative, dispositional, and recipient-passive meanings (Panevová et al., in print). The surface structure of a verb with active voice is considered to be the unmarked member of a diathesis, whereas the structure with the given verb characterized by some of the five above given meanings constitutes its marked member. Whereas the passive, resultative and recipient-passive meanings of verbs are formed by the auxiliary verbs *být* (passive and resultative d.), *mít* (resultative d.), and *dostat* (recipient-passive d.), respectively, plus past participle of a lexical verb, the deagentive and dispositional diatheses are associated with the reflexive verb form. This form is constituted by the active form of a verb and the freestanding morpheme *se*.

4.1 Deagentive Diathesis

Marked members of the deagentive diathesis prototypically imply agentive ACTor of the event expressed by the verb; however, the ACTor is never expressed in the surface structure. The use of the deagentive meaning of a verb results in specific changes in its valency structure. In Czech, deagentive meaning can be applied to both transitive and intransitive verbs; the reflexive verb form is limited to 3rd person.⁷ For a transitive verb, (i) ACTor is shifted from the subject position (expressed in the nominative) and (ii) the subject is filled with the argument of the verb corresponding to the direct object in the unmarked construction (prototypically an accusative object), as in (26). For an intransitive verb, the shift of ACTor away from the subject position results in a subjectless surface structure in which the verb has prototypical form of 3rd sg neutrum, as in (27).

(26a) *Dělníci opravují silnici.*

“workers_{ACT-Subj-nom} – repair – the road_{PAT-Dobj-acc.}”
Eng. Workers repair the road.

(26b) *Silnice se opravuje.*

“the road_{PAT-Subj-nom} – SE_{morph} – repairs.”
Eng. The road is being repaired.

(27a) *Lidé na večírku tančili.*

“People_{ACT-Subj-nom} – at the party – danced.”
Eng. People danced at the party.

(27b) *Na večírku se tančilo.*

“at the party – SE_{morph} – danced_{3rd-sg-neutr.}”
Eng. At the party, there was some dancing.

4.2 Dispositional Diathesis

As in the case of the deagentive diathesis, the marked members of the dispositional diathesis indicate human ACTor that is shifted from the subject position (in the nominative). This position is filled by the argument corresponding to the direct object position of a transitive verb, see example (28); in the case of an intransitive verb, dispositional meaning of the verb results in a subjectless surface structure (with the 3rd sg neutrum verb form), see example (29). In contrast to the deagentive diathesis, ACTor can be optionally expressed as an indirect object expressed by the dative. The marked members of dispositional diathesis are characterized by the presence of evaluative adverbs; thus if ACTor is expressed in the surface structure, it can be interpreted as an evaluator, see examples (28) and (29).

(28a) *Petr četl tuto knihu.*

“Peter_{ACT-Subj-nom} – read – this book_{PAT-Obj-acc.}”

⁷ The status of several constructions with 2nd person (i) and 1st person (ii) is rather unclear, as they can be interpreted as either deagentive (with the grammatical morpheme *se*) (i), or reflexive (with the reflexive pronoun *se*) (ii).

(i) *Odsuzujete se k pěti letům vězení.*

“Sentence_{2nd-pl-masc/fem} – SE_{morph/pron} – to five year's imprisonment.”

Eng. You are sentenced to five year's imprisonment. // You sentence yourself to five year's imprisonment.

(ii) *Léčím se u doktora Nováka.*

“Treat_{1st-sg} – SE_{morph/pron} – at the doctor Novák.”

Eng. I am treated at doctor Novák. // I treat myself by doctor Novák.

With respect to the low frequency and lexical constraints on the lexical expression of arguments, these Czech constructions can be regarded as close to idioms. For this reason, we leave their representation in the lexicon aside.

Eng. Peter read this book.

(28b) *Tato kniha se (Petrovi) dobře četla.*

“this book_{PAT-Subj-nom} – SE_{morph} – (Peter_{ACT-IObj-dat}) – well – read.”

Eng. This book read well.

(29a) *Já jsem tam spal.*

“I_{ACT-Subj-nom} – am – there – slept.”

Eng. I slept there.

(29b) *Spalo se (mi) tam dobře.*

“slept_{3rd-sg-neutr} – SE_{morph} – (me_{ACT-IObj-dat}) – there – well.”

Eng. I slept well there.

4.3 Representation in the Lexicon

The changes in the valency structure of verbs in both deagentive and dispositional diatheses involve changes in morphemic forms of the arguments affected by surface shifts. These changes are regular enough to be captured by formal rules, which are stored in the grammar component of the lexicon. In the data component, only valency frames corresponding to the unmarked (active) uses of a verb are recorded. The optional attribute -diat is attached to each relevant valency frame in the data component (see Figure 1); it provides the information on applicability of the specific morphological meaning(s). On the basis of the formal rules (see Figure 2), the valency frames corresponding to marked structures of diatheses can be automatically derived (Kettnerová et al., 2012b).

| |
|---|
| <p>- lemma: <i>zabíjet</i>^{impf}, <i>zabít</i>^{pf} 'to kill' - gloss: impf: <i>usmrcovat</i> pf: <i>usmrtit</i> 'to cause death' - frame: ACT₁^{obl} PAT₄^{obl} - example: impf: <i>zabíjet někoho nožem; zabíjeli mi manžela před očima</i> pf: <i>zabít někoho nožem; zabili mi manžela před očima</i> 'to kill sb with a knife; they killed my husband before my eyes' - rfl: cor4: impf: <i>zabíjel se z nešťastné lásky několikrát do roka</i> 'he used to kill himself several times a year' pf: <i>zabil se z nešťastné lásky</i> 'he killed himself' - rcp: ACT-PAT: impf: <i>zabíjeli se navzájem</i> pf: <i>zabili se navzájem</i> 'they killed each other' - diat: Deagent: impf: <i>o masopustu se každoročně zabíjí prase</i> pf: <i>o masopustu se zabilo</i> 'during the carnival a pig is killed yearly' Disp: impf: <i>prase se řezníkovi špatně zabíjelo</i> 'the pig killed bad'</p> |
| <p>- lemma: <i>zabíjet</i>^{impf} <i>se</i>, <i>zabít</i>^{pf} <i>se</i> 'to kill oneself, to die' - gloss: impf: <i>umírat; usmrcovat se</i> pf: <i>zemřít; usmrtit se</i> 'to die' - frame: ACT₁^{obl} - example: impf: <i>každý rok se na hřištích zabije několik dětí</i> 'each year several children die (by accident) at the playground' pf: <i>zabil se na kole při havárii</i> 'he died during the car accident with his bicycle'</p> |

Figure 1: Example of two lexical units of two lexemes *zabíjet*^{impf}, *zabít*^{pf} “to kill” and *zabíjet*^{impf} *se*, *zabít*^{pf} *se* “to kill oneself, to die” in the data component of the Valency Lexicon of Czech Verbs, VALLEX.

5 Conclusion

We have discussed the possibility of a lexicographic representation of Czech reflexive verbs. We have shown that reflexivity should be described in different ways, depending on the function of the reflexive morphemes *se* and *si*: (i) Reflexive tantum verbs and derived reflexive verbs (where *se/si* is a part of a verb lemma) are stored as separate verb lexemes (represented by separate verb lemmas) in the data component of the lexicon. (ii) The possibility of a verb to be used in reflexive constructions (in the narrow sense) and in reciprocal constructions (where *se* is the personal pronoun coreferring to the subject) is marked in the special attributes -rfl and -rcp, respectively, assigned to relevant lexical units in the data component. For reciprocal constructions, formal rules stored in the grammar component make it possible to automatically derive respective valency frames. (iii) Similarly, the possibility of a verb to undergo deagentive or

dispositional diathesis (where *se* is part of the verb form) is marked in the attribute -diat assigned to each relevant lexical unit in the data component; formal rules stored in the grammar component enable the derivation of the valency frames underlying the marked members of the diatheses.

| Type: Deagent | | | Commentary |
|------------------|----------|------------------------------------|------------|
| Action | verbform | replace (active vf → reflexive vf) | (1) |
| | ACT | delete (nom → ∅) | (2) |
| | PAT | replace (acc → nom) | (3) |

Commentary:

(1) The verb form changes from active to reflexive (adding the reflexive *se*).

(2) ACTor cannot be expressed in the surface syntactic structure.

(3) The morphemic expression of PATient changes from the accusative into thenominative (and shifts to the subject position).

Figure 2: Example of a formal rule in the grammar component of the Valency Lexicon of Czech Verbs, VALLEX – the rule describing the deagentive diathesis.

6 References

- Dokulil, M. (1986). *Mluvnice češtiny I*. Praha: Academia.
- Hajič, J., Panevová, J., Urešová, Z., Bémová, A., Kolářová, V. & Pajas, P. (2003). PDT-VALLEX: Creating a Large-Coverage Valency Lexicon for Treebank Annotation. In Proceedings of the Second Workshop on Treebanks and Linguistic Theories, 15-15 November. Vaxjö, Sweden, pp. 57-68.
- Hajič, J., Hajičová, E., Panevová, J., Sgall, P., Pajas, P., Štěpánek, J., Havelka, J. & Mikulová, M. (2006). Prague Dependency Treebank 2.0. Philadelphia, PA: Linguistic Data Consortium. LDC2006T01.
- Hajičová, E., Panevová, J. & Sgall, P. (1985,1986,1987). Coreference in the Grammar and in the Text. Part I. In *The Prague Bulletin of Mathematical Linguistics*, 44, pp. 3-22. Part II. In *The Prague Bulletin of Mathematical Linguistics*, 46, pp. 1-11. Part III. In *The Prague Bulletin of Mathematical Linguistics*, 48, pp. 3-12.
- Kettnerová, V., Lopatková, M. & Bejček, E. (2012a). The Syntax-Semantics Interface of Czech Verbs in the Valency Lexicon. In *Proceedings of the 15th Euralex International Congress 2012, 7-11 August 2012*. University of Oslo, Norway, pp. 434-443.
- Kettnerová, V., Lopatková, M. & Urešová, Z. (2012b). The Rule-Based Approach to Czech Grammaticalized Alternations. In *Proceedings of the 15th International Conference Text, Speech, Dialogue 2012, 3-7 September 2012*. Masaryk University, Czech republic, pp. 158-165.
- Lenci, A., Lapesa, G. & Bonansinga, G. (2012) LexIt: A Computational Resource on Italian Argument Structure. In *Proceedings of LREC 2012*, pp. 3712- 3718.
- Lopatková, M., Žabokrtský, Z., & Kettnerová, V. (2008). *Valenční slovník českých sloves*. Praha: Nakladatelství Karolinum.
- König, E., Gast, V. (2008). *Reciprocals and Reflexives: Theoretical and Typological Explorations*. Berlin, New York: Mouton de Gruyter.
- Nedjalkov, V. (2007). *Typology of Reciprocal Constructions*. Amsterdam: Benjamins.
- Oliva, K. (2001). Reflexe reflexivity reflexive. In *Slovo a slovesnost*, 57, pp. 200-207.
- Oliva, K. (2003). Linguistics-based PoS-tagging of Czech: disambiguation of *se* as a test. In *Contributions of the 4th European Conference on Formal Description of Slavic Languages, 28-30 November 2001*. Postdam University, Germany, pp. 299-314.
- Panevová, J. (1994). Valency Frames and the Meaning of the Sentence. In P.A. Luelsdorff (ed.) *The Prague School of Structural and Functional Linguistics*. Amsterdam, Philadelphia: John Benjamins Publishing Company, pp. 223-243.
- Panevová, J. (1999). Česká reciproční zájmena a slovesná valence. In *Slovo a slovesnost*, 60, pp. 269-275.
- Panevová, J. (2007). Znovu o reciprocitě. In *Slovo a slovesnost*, 68, pp. 91-100.
- Panevová, J., Mikulová, M. (2007). On Reciprocity. In *The Prague Bulletin of Mathematical Linguistics*, 87, pp. 27-40.
- Panevová et al. (in print). *Mluvnice současné češtiny. Část 2: Syntax češtiny na základě anotovaného korpusu*. Praha: Nakladatelství Karolinum.
- Petkevič, V. (2013). Formal (Morpho)Syntactic Properties of Reflexive Particles *se*, *si* as Free Morphemes in Contemporary Czech. In *Proceedings of the 7th International Conference 2013, 13-15 November 2013*. Slovenská akadémia vied, Slovakia, pp. 206-216.
- Renau, I., Battaner, P. (2012). Using CPA to Represent Spanish pronominal Verbs in a Learner's Dictionary. In *Proceedings of the 15th Euralex International Congress 2012, 7-11 August 2012*. University of Oslo, Norway, pp. 350-361.
- Ruppenhofer, J., Ellsworth, M., Petruck, M.R.L., Johnson, Ch.R. & Scheffczyk, J. (2010) *FrameNet II: Extended Theory and Practice*. <https://framenet2.icsi.berkeley.edu/docs/r1.5/book.pdf>
- Sgall, P., Hajičová, E. & Panevová, J. (1986). *The Meaning of the Sentence in Its Pragmatic and Semantic Aspects*. Dordrecht: Reidel.

Skoumalová, H. (2001). *Czech syntactic lexicon*. PhD thesis. Charles University, Prague, Czech Republic.
Žabokrtský, Z., Lopatková, M. (2007) Valency Information in VALLEX 2.0: Logical Structure of the Lexicon. In *The Prague Bulletin of Mathematical Linguistics*, 87, pp. 41-60.

Acknowledgements

The research reported in this paper has been supported by the Czech Science Foundation GA ČR, grant No. P406/12/0557. This work has been using language resources developed and/or stored and/or distributed by the LINDAT/CLARIN project of the Ministry of Education, Youth and Sports of the Czech Republic (project LM2010013).