Univerzita Karlova v Praze
Filozofická fakulta
Ústav teoretické a komputační lingvistiky

Czech syntactic lexicon

Hana Skoumalová

2001

filologie — matematická lingvistika
vedoucí práce — prof. PhDr. Jarmila Panevová, DrSc.

Acknowledgments

I would like to thank Prof. Jarmila Panevová, for her tutoring. She introduced me to the valency theory, she led me in my work, and she pointed out work to me that has been done by generations of linguists when I tried to reinvent the wheel.

I also want to thank my colleagues at work, Niki Petkevič, Saša Rosen and Milena Hnátková who were exposed to my crazy ideas and who helped me to judge some strange sentences and syntactic constructions.

My gratitude belongs to my parents—my mother Zdena Skoumalová and my late father Jan Skoumal—who trained me in linguistics since my childhood. It was my father who showed me that linguistics is fun, and my mother who showed me that linguistics is serious work. In spite of their dissuasion, I continue the family tradition and I hope that I will not thereby harm it.

My thanks go also to Karel Oliva who read a part of the dissertation and gave me several useful remarks, and Clare Britton who corrected my English.

Last but not least, I must thank the Research Support Scheme, which supported my work with grant no. 72/94, the Grant Agency of the Czech Republic, which supported the work by grant no. 405/96/K214, and the Research and Development Council of the CR Government, which supported the work by Research program MSM 112100002.

Contents

List of Tables

List of Figures

1. Introduction

In the era of computers, language processing has gained a form different from what was known before. Vast amounts of data are available and computers can process them in a reasonably short time, but they need adequate tools for their work. Beside grammar rules they also need lexicons which they can understand.

In this work, an electronic lexicon of Czech verbs is presented. The use of the lexicon in Natural Language Processing (NLP) makes special demands on it. It differs from "human" lexicons in that all information must be explicit or deducible by exactly formulated rules of derivation.

While sketching the format of the dictionary, interesting theoretical problems were encountered, which are discussed in this work. Though the lexicon should not depend heavily on a particular theory, so that it can remain usable in another theoretical frame, it is impossible to make it totally theory-free. It is possible, however, to design the dictionary in such a manner that it will not be difficult to adapt it for a particular theoretical frame. The possibility to reuse our lexicon in other frameworks will be discussed at the end of the work.

The lexicon contains valency frames of ca 15,000 Czech verbs, and its purpose is to enrich information contained in other electronic dictionaries. The trend of recent years is to make large-scale reusable sources which can be combined with other sources. This work shows how the lexicon cooperates with an existing morphological lexicon and how it can be used in various NLP systems.

Chapter 2 discusses several theoretical approaches in comparison with Functional Generative Description (FGD), which is used for the dictionary. The explication concentrates especially on the structure of lexicons in single theories. A lexicon usually conforms certain preconditions resulting from using a given theoretical framework and we will explore the possibility of creating a lexicon which would be transferable to another theoretical framework.

Chapter 3 discusses the possibility of using existing sources, with respect to the desired result and the theoretical framework adopted for the work. There were already several Czech syntactic lexicons created in the past, but unfortunately their reuse would be rather difficult. This chapter mentions several such attempts, and describes in detail a lexicon which was used.

Chapter 4 describes the verb frame. In **Section 4.1** we will describe the format of the lexical entry. In **Section 4.2** we will discuss various types of reflexive constructions in Czech, and their encoding in the lexicon. In **Section 4.3**, possible diatheses of the basic (active) frame are shown, and it is also discussed which of these diatheses can be added to the dictionary on a regular basis and which have to be treated as exceptions. **Section 4.4** describes so called equi and raising verbs.

In **Chapter 5** we will show the procedure of automatic conversion of the source dictionary to the proposed format. For this conversion, an algorithm was created which assigns the functors (semantic roles) to single members of a frame. The output of this procedure will serve as an input for an editor. We will discuss what amount of the source data can be completed by this procedure and what amount needs post-editing. We will also show how the resulting lexicon can be used in NLP systems.

Chapter 6 sums up. In **Section 6.1**, verbs are sorted into groups according their frames, and the results are compared with results of other researchers. In **Section 6.2**, perspectives of the language processing based on symbolic methods are discussed, and the possible usage of the lexicon in corpus linguistics.

1.1. Terminological remarks

Various authors differ their in understanding of the term *subject*. We will consider a subject only such a member of a frame which is in Nominative and with which the main verb agrees. Our criterion is the question for a subject: *kdo, co* (who$_{Nom}$, what$_{Nom}$). This means that we will not take Genitive in such constructions as *vody ubývá* (water$_{Gen}$diminishes) as subject. On the other hand, a clause or infinitive can be subjects, as we can ask the above question; in such a case the verb shows ageement with neuter singular.

In the text, we will use the terms *actants* and *inner participants* as synonymous. *Actant* is Tesnière's term, while FGD uses *inner participant*, but their meaning is so close that they are often interchanged.

We will also use the terms *animate* and *animacy*. For purposes of this work we will divide nouns into two groups: personal and non-personal. The former can be represented by the pronoun *kdo* (who), the latter by the pronoun *co* (what). Sometimes we will refer to personal nouns as to animate ones and to non-personal as to inanimate.

2. Theoretical background

When describing the role of verbs in the language, all authors agree on the necessity to describe syntactic properties of verbs in the dictionary. But they differ in the understanding of what sort of information should be included. Dictionaries for practical usage (language dictionaries for human readers, or machine dictionaries for grammar checking or shallow parsing) contain usually only the surface information.

Dictionaries that serve more sophisticated purposes must contain also information on the argument structure, and the relations between the two layers of linguistic description. The two views of the dictionary differ also in their understanding of what belongs to the verb frame. The classical lexicologists collect all *typical* complementations while the theoreticians discriminate between the *arguments* and *adjuncts*. The arguments are listed in subcat lists and grammar rules check whether all of them are present in a sentence. The adjuncts, on the other hand, are not obligatory, can occur more than once in a sentence, and they are not listed in the dictionary entries.

The dictionary described in this work is meant to provide for the automatic processing of the Czech language. The algorithms for the language processing do not necessarily have to be based on a linguistic theory, but we believe that with a theory we can develop algorithms that are efficient and elegant because they are linguistically adequate. The results of these algorithms, on the other hand, can serve to a linguistic theory as a feedback which helps to improve the theoretical description.

For this work we decided to utilize the Functional Generative Description (FGD) developed by Sgall, Hajičová and Panevová (Sgall et al., 1986), and especially the part dealing with the verb frames (Panevová, 4 75, 1980). We will show later that this does not prevent the dictionary from being used in other theoretical frameworks.

2.1. An overview of FGD

In FGD, several levels of language description are distinguished. For purposes of this work, we will only work with two of them—the *tectogrammatical* level and the *morphemic* level. To be able to express certain relations we will also need the notion of *subject*.

Each level has its own units, basic and compound. The compound units are formed from the basic ones with the help of *C-relations*. The translation between two neighbouring levels is provided by *R-relations*. The basic units on tectogrammatical level are *semantemes* (lexical units), *functors* (syntactic units) and *grammatemes*. The compound units are *propositions*. The functors also serve as the C-relations with the help of which the propositions are constructed (see Sgall, 1967).

There are two types of functors—*inner participants* (Tesnière's *actants*) and *free modifications*. A verb frame denotes which functors are required by a certain semanteme (verb lemma). A frame can contain up to five inner participants (*Actor*, *Patient*, *Addressee*, *Origin* and *Effect*) and any number of free modifications. Some of the inner participants can be *optional* (also called *facultative*), which means that they do not need to be present in the sentence—neither on the tectogrammatical nor morphemic level. Other participants are always *obligatory*. However, they can be realized as *general*—the structure on the tectogramatical level then contains a general participant, which is not realized on the morphemic level. Whether a participant is optional or obligatory, and whether an obligatory participant can be realized as general can be tested by a question test (Panevová, 1980, pp.29-30). Let us imagine the following dialogue:

(1) *Petr čte.* *Co?* *Nevím.*
Petr is reading. What? I don't know.

The answer 'I don't know' is acceptable, as the the speaker does not need to know what Petr's reading is, but it must be something which is usually read (a newspaper, a book, etc). This shows that Patient in the frame of the verb *číst* (read) can be general. On the other hand, in dialogue (2), the answer 'I don't know' is nonsensical. This shows that Actor is an obligatory participant in the frame of the verb *přijít* (come).[1]

(2) *Už přišel.* *Kdo?* *Nevím.*
(He) has already come. Who? I don't know.

In example (3) the sentence is actually ungrammatical, if the participant is omitted—this is clear evidence that the participant is obligatory.

(3) **Petr daroval.*
Petr donated.

Free modifications normally are not members of a frame, but they can become members as *obligatory free modifications*:

(4) a. *Jan se choval jako blázen.*
Jan behaved like a fool.

[1]The fact that the surface realization of Actor in this sentence is omitted is caused by another phenomenon: Czech is a so called pro-drop language and thus a personal pronoun in the position of a subject can be omitted. Morphological markers of the person and number (in the past tense also of the gender) are present also in the verb form and thus the personal pronoun is redundant (see Karlík, 2000).

b. **Jan se choval.*
Jan behaved.

In some cases, when the modification is known from the context, it can be omitted on the surface; such free modification is called *obligatory and deletable free modification*. For testing whether a free modification is an obligatory member of a frame the question test can be used again. In the sentence in (5) the question test proves that the direction is an obligatory and deletable free modification of the verb *přijít* (come, arrive).

(5) *Petr přišel. Kam? *Nevím.*
Petr arrived. Where? I don't know.

In other theoretical models (Daneš et al., 1987a; Grepl and Karlík, 1989; Karlík et al., 1995), the repertory of participants is wider: instead of Actor the authors speak about Agent, Causer, Experiencer, etc. Patient is more or less a synonym of the direct object and Recipient a synonym of the indirect object. In FGD, Actor and Patient are determined by syntactic criteria rather than by semantic ones (cf. Tesnière, 1959), and other participants are determined semantically:

(6)
1. If the verb frame contains only one participant, this participant is Actor.
2. If the frame contains two participants, one of them is Actor and the other is Patient. In most cases, Actor is the subject of the active construction, but there are some exceptions to this rule, which will be discussed later.
3. If the verb frame has more than two participants, the roles of Actor and Patient must be occupied, and the other participants occupy the roles of Addressee, Effect or Origin. The decision about which participant bears which role is based on the semantics of the participants.

The basic units on the morphemic level are *semata*, and the compound units are *morphemes* and *formemes*—units which combine prepositions with morphological cases.

The lexicon in FGD contains semantemes, their functors and grammatemes. In our informal example, parantheses denote whether a functor is obligatory, obligatory deletable or optional:

(7)

spát	Act		
pojídat	Act	Pat	
těšit_se	Act	Pat	Gram:{Refl[se]}
darovat	Act	Pat	(Addr)

Beside it, the lexicon should also define the R-relation which translates every functor and grammateme to the morphemic level. After this addition, the lexicon will have the following format:

(8)	*spát*	Act[Noun+Nom]		
	pojídat	Act[Noun+Nom]	Pat[Noun+Acc]	
	těšit_se	Act[Noun+Nom]	Pat[Noun+Acc+na]	Gram:{Refl[se]}
	darovat	Act[Noun+Nom]	Pat[Noun+Acc]	(Addr[Noun+Dat])

2.2. Comparing FGD with other theories

In this section, a short comparison of the main contemporary linguistic theories is provided and the possibility of interchange of a common dictionary is discussed.

2.2.1. Government-Binding Theory

In Government-Binding Theory, the lexicon contains words with subcat lists and lists of θ-roles. The match between arguments of a verb and θ-roles is taken care of by θ-Criterion:

> Each argument bears one and only one θ-role, and each θ-role is assigned to one and only one argument.

The match between categories in a subcat list and θ-roles is called θ-marking:

> If α subcategorizes the position occupied by β, then α θ-marks β.

The lexicon in GB then contains the *word*, its *category*, *subcat list* and list of θ roles:

(9) sneeze, V, (<u>Agent</u>)
devour, V, <NP>, (<u>Agent</u>, Theme)
donate, V, <NP, PP>, (<u>Agent</u>, Theme, Goal)

where the θ-roles are results of θ-marking.

Subjects do not occur in subcat lists, as it is presupposed that every verb has a subject. The theoretical explanation is that the subject is an *external* argument.

Passivization in GB is provided by movement rules. Active sentences are transformed to other constructions and ungrammatical structures are then ruled out by various principles that exploit θ-roles assignment to single complementations.

2.2.2. Lexical-Functional Grammar

In Lexical-Functional Grammar, the dictionary plays the central role (as the name suggests). The theory works with grammatical categories (as NP, S′, XCOMP, etc.) and grammatical functions (as Subject, Object, etc.). Categories are used for constructing c-structures, while functions are used for f-structures. The lexicon in LFG has the following format:

(10) *sneeze* V ($\uparrow$PRED) = 'sneeze<($\uparrow$SUBJ)>'
devour V ($\uparrow$PRED) = 'devour<($\uparrow$SUBJ), ($\uparrow$OBJ)>'
donate V ($\uparrow$PRED) = 'donate<($\uparrow$SUBJ), ($\uparrow$OBJ), ($\uparrow$OBL$_{GO}$)>'

The theory does not work with *arguments* directly, but it supposes some sort of linking between θ-roles and grammatical functions called *Predicate-Argument Structure*. An enhancement of the theory is the *semantic structure* which, however, works with concepts as ARG$_1$, ARG$_2$, rather than with θ-roles.

Frames for passive sentences are created with the help of lexical rules which may have the following form:

(11) ($\uparrow$SUBJ) $\mapsto$ NULL
($\uparrow$OBJ) $\mapsto$ ($\uparrow$SUBJ)
~($\uparrow$TENSE)
($\uparrow$PARTICLE)$=_c$ PASS

These rules erase the original subject from the frame, move the object to its place, and they add two constraints on the verb form—it must not be a finite form and the value of the attribute PARTICLE must be PASS.

2.2.3. Head-Driven Phrase Structure Grammar

HPSG works with *signs* which are in fact *well typed* attribute value matrices (AVM).[2] The whole grammar is based on combining AVM's together with the help of unification. A lexical entry has a form of an AVM, too:

(12) *walks*

$$\begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{HEAD} & {}_{verb}\begin{bmatrix} fin \end{bmatrix} \\ \text{SUBCAT} & \left\langle \text{NP}[nom]_{\boxed{1}[3rd,sing]} \right\rangle \end{bmatrix} \\ \text{CONTENT} & \begin{bmatrix} \text{RELN} & walk \\ \text{WALKER} & \boxed{1} \end{bmatrix} \end{bmatrix}$$

[2] The term *well typed AVM* means that what attributes can appear in an AVM is determined by its type.

sees $$\begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{HEAD} & {}_{verb}\begin{bmatrix} fin \end{bmatrix} \\ \text{SUBCAT} & \left\langle \text{NP}[nom]_{\boxed{1}[3rd,sing]},\ \text{NP}[acc]_{\boxed{2}} \right\rangle \end{bmatrix} \\ \text{CONTENT} & \begin{bmatrix} \text{RELN} & see \\ \text{SEER} & \boxed{1} \\ \text{SEEN} & \boxed{2} \end{bmatrix} \end{bmatrix}$$

gives $$\begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{HEAD} & {}_{verb}\begin{bmatrix} fin \end{bmatrix} \\ \text{SUBCAT} & \left\langle \text{NP}[nom]_{\boxed{1}[3rd,sing]},\ \text{NP}[acc]_{\boxed{2}},\ \text{NP}[acc]_{\boxed{3}} \right\rangle \end{bmatrix} \\ \text{CONTENT} & \begin{bmatrix} \text{RELN} & give \\ \text{GIVER} & \boxed{1} \\ \text{GIVEN} & \boxed{2} \\ \text{GIFT} & \boxed{3} \end{bmatrix} \end{bmatrix}$$

The valency frame is contained in the attribute SUBCAT. A mapping between the subcat list and CONTENT is provided by the indices ($\boxed{1}$, $\boxed{2}$, $\boxed{3}$, etc). The attributes in CONTENT are not marked as ARG_1, ARG_2, etc., as one would expect but their names are derived from the verb lemma. For linking the arguments with θ-roles, so called *linking theory* is used.

Passive frames are created with the help of lexical rules. They change the chracteristics of the verb form and cyclically permute subcat lists, as shown in (13):

(13) SUBCAT $<NP_1,NP_2> \mapsto$ SUBCAT $<NP_2,PP[by]_1>$
SUBCAT $<NP_1,NP_2,NP_3> \mapsto$ SUBCAT $<NP_2,NP_3,PP[by]_1>$
SUBCAT $<NP_1,NP_2,PP[to]_3> \mapsto$ SUBCAT $<NP_2,PP[to]_3,PP[by]_1>$
...

The resulting lexical entries then look as shown in (14):

(14) *seen* $$\begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{HEAD} & {}_{verb}\begin{bmatrix} pass \end{bmatrix} \\ \text{SUBCAT} & \left\langle \text{NP}[nom]_{\boxed{2}[3rd,sing]},\ \text{PP}[by]_{\boxed{1}} \right\rangle \end{bmatrix} \\ \text{CONTENT} & \begin{bmatrix} \text{RELN} & see \\ \text{SEER} & \boxed{1} \\ \text{SEEN} & \boxed{2} \end{bmatrix} \end{bmatrix}$$

$$
\textit{given} \quad \left[\begin{array}{ll}
\text{CAT} & \left[\begin{array}{ll}
\text{HEAD} & {}_{verb}\left[\textit{pass}\right] \\
\text{SUBCAT} & \left\langle \text{NP}[nom]_{\boxed{2}[3rd,sing]},\ \text{NP}[acc]_{\boxed{3}},\ \text{PP}[by]_{\boxed{1}} \right\rangle
\end{array}\right] \\
\text{CONTENT} & \left[\begin{array}{ll}
\text{RELN} & \textit{give} \\
\text{GIVER} & \boxed{1} \\
\text{GIVEN} & \boxed{2} \\
\text{GIFT} & \boxed{3}
\end{array}\right]
\end{array}\right]
$$

Some authors argue (Oliva, 1994; Kathol, 1994) that lexical rules are not necessary, as the desired effect could be achived by applying constraints on the hierarchy of types, but we will not go to details here.

2.2.4. Comparison with FGD

In all the above mentioned theories, some sort of mapping between surface forms and θ-roles is supposed, whether it is called θ-marking, predicate-argument structure, or linking theory. The common feature is that subcat lists are viewed as primary syntactic structure attached to lexical entries and the θ-roles are mapped onto the subcat list by some sort of mapping function.

In FGD the opposite assumption is made: the tectogrammatical functors form a primary syntactic structure of a verb and the surface forms are their counterparts on the morphemic level which are translated by R-relation from the functors.

Beside this, the θ-roles differ from the repertory of participants in FGD. Not only are their names different, but also their distributions to single verbs. An *Actor* in FGD can be marked as *Agent* or *Bearer* or *Experiencer* in other theories, etc.

If we use FGD as the background theory of a dictionary, we will be unable to transfer the lexicon to another theoretical framework ‘as is’; it should not be difficult, however, to extract the subcat lists. It will be shown in Chapter 6 that this is possible and feasible. For utilizing the tectogrammatical information, we would have to find a mapping function which would have to take into consideration also the semantics of single verbs, which will be the subject of further research.

3. Using existing sources

When we try to create a new electronic dictionary, it is of course possible to start from scratch, but it is more efficient to use existing sources. Printed dictionaries usually contain syntactic information, but unfortunately this information is meant for human readers, and very often it is assumed that the reader knows the rules that apply in usual cases, and only exceptions are listed. Beside this, the information is not encoded in a formal way which could be understandable to a machine.

There exists a Czech dictionary of verbs (see Svozilová et al., 1997) which contains the verb frames encoded in a formal way. But its size is quite limited (ca 600 verbs) and the information concerns only the surface frames. Nevertheless, this dictionary can serve as an aid to creators of an electronic dictionary.

One of the first attempts at making an electronic dictionary of verb frames was made in the project RUSLAN (see Oliva, 1989). This project was focused on machine translation from Czech to Russian and the format of the lexicon was adapted for this purpose; it contained the Czech word stem and its Russian translation, Czech and Russian morphological information, the Czech surface frame and its translation to the Russian surface frame. The domain of the translated texts were programming manuals, which affected the coverage of the lexicon. Another drawback (caused by limited computational resources) was the small size of the lexicon—it contained ca 10,000 entries (including all word classes). The work invested in this project was useful for gaining experience with natural language processing rather than for creating working software.

Another small lexicon was created for the purposes of the project LaTeSlav (see Avgustinova et al., 1995). This was a project for creating grammar-checkers for two Slavic languages (Czech and Bulgarian). In fact, there were two lexicons for Czech, as the project split into two branches. Both the lexicons (see Oliva, 1996; Skoumalová, 1994) contained a small number of entries which had very rich syntactic information, but unfortunately they were "hardwired" in the software and it would not be easy to extract them for other purposes.

The most promising source of valency frames is a dictionary created at Masaryk University by Karel Pala and his team (see Pala and Ševeček, 1997; Horák, 1998b). This dictionary was compiled from several printed dictionaries, and the valency frames were taken mainly from SSJČ (1989). We used this dictionary as a source of surface frames and enhanced them with information at the tectogrammatical level.

3.1. Source data

The dictionary contains ca 15,000 verbs with surface frames. The original format called BRIEF contains lemma, starting delimiter of the list of frames (<v>) and the list itself (see example in 15a). (15b) translates this notation to a readable form.

(15) a. `agitovat <v>hPc4,hPc3-hPc4,hPTc4r{pro},hPTc3r{proti}`

b. **agitovat** (to agitate) *koho* (*komu*), pro *koho*, proti *komu*

In BRIEF format, frames are separated by commas, and single members of a frame are separated by dashes. The obligatoriness is not marked, but a frame can be repeated several times, with and without the optional, deletable or generalizable members. In example (15) this is the case of the frame *koho* (*komu*).

BRIEF encoding is described in Horák (1998a,b). Here, we only provide a short overview of attributes and values used in the dictionary. Every member of a frame is described by a list of attributes and their values. We can understand these attributes and their values as grammatemes occurring on the tectogrammatical level.

3.1.1. The attributes used in the lexicon and their values

`h` — 'Semantic' feature. This attribute has rather heterogeneous values. Single values are only applicable for certain word classes and thus they include implicit information on the part of speech as well. The values are:

- `P` — Person (only applicable for nouns and pronouns); this value actually stands for 'case questions' *kdo* (who), *koho*, etc.
- `T` — Thing (only nouns an pronouns); it stands for 'case questions' *co* (what), *čemu*, etc. The values P and T can be grouped together.
- `R` — Long reflexive pronoun *sebe*, *sobě*, etc.
- `Q` — Quality (only adjectives).
- `M` — Amount (only numbers).
- `L` — Location (only adverbs).
- `A` — Direction where (only adverbs).
- `F` — Direction from (only adverbs).
- `D` — Which way (only adverbs).
- `W` — When (only adverbs).

`c` — Morphological case. This attribute is only applicable for nominal word classes, and so it only occurs if the h attribute has one of the values P, T, R, or Q. The values are 1, 2, 3, 4, 6 and 7.

`r` — Preposition. This attribute can only occur after a morphological case. The value is the preposition itself closed in curly brackets: r{na}, r{o}, r{vzhledem k}, etc.

`s` — Clause or infinitive. The values are:

- `I` — Infinitive.
- `C` — Clause attached by the conjunction *až* (when).
- `D` — Clause attached by the conjunction *že* (that).
- `F` — Clause attached by the conjunction *jestli*, *zda* (if, whether).
- `P` — Clause attached by the conjunction *ať* (let).
- `R` — Clause attached by a relative expression *co* (what), *který* (which), *kdo* (who), *kolik* (how many), etc.
- `U` — Clause attached by the conjunction *aby* (so that).
- `Z` — Clause attached by the conjunction *jak* (how).

`e` — Negation (in a clause). The values are A (affirmative) and N (negative). The affirmative value is the default and it is not marked in the lexicon.[1]

`i` — Idiom. The value is a string closed in curly brackets. The string contains words forming the idiom and a case marker for the variable part. If there are possible variants in the fixed part, they are put in parentheses and separated by commas, or they are separated by a vertical bar. The variants in the variable part are separated by a vertical bar. Examples:

```
brát <v>i{pod ochranu|do ochrany <koho>}
                                        (place sb under protection)
dávat <v>i{koncert|hru|film}           (put concert, play, movie on)
házet <v>i{přes palubu <koho|co>}                 (throw sb over board)
chovat <v>i{(přátelství, zášť, nenávist) <ke komu>}
                                          (feel friendship, hatred)
```

`v` — Constraint applied for a single valency frame. The constraint is an attribute with a required value, or an attribute with a prohibited value, preceded by ^. Currently, only v{eN} is used, for verbs whose negated forms have different valency frames:

```
hledět <v>hPTc4r{na},hPc3,hTc2r{do},hPc3-hTc2r{do},v{eN}hTc3r{k}
                                                 (not to look at st)
```

[1] This attribute is mainly used together with a clause attached by the conjunction *aby* (so that)—sUeN, e.g. *bát se* (fear), *varovat* (warn), etc. Though this is a typical usage, the affirmative clause cannot be excluded. After a simple query in the Czech National Corpus (Kocek et al., 2000) we found eight affirmative clauses (out of ca 230 occurrences of the verb *bát se* with the conjunction *aby*), e.g. *Po volbách se úředníci bojí, aby přežili ... změnu dnešního ministra ...* (After elections, clerks are afraid whether they will survive the change of the current minister ...).

```
páchnout <v>hTc6r{po},hTc7,v{eN}hTc2r{do} (not to set foot on st)
znát <v>hPTc4,v{eN}hTc2z{jen se záporem},hTc4-hPTc6r{na}
                                      (not to know--Genitive of negation)
```

`z` — Comment in curly brackets (see the example above).

The frames do not contain subjects as the printed dictionaries usually do not list them. For an automatic processing of language, however, this information is necessary. We can make a simple assumption that the subject will be a noun in Nominative but there are exceptions to this rule. We will discuss this in more detail in Chapter 5.

4. Content of the lexicon

In this chapter, a detailed description of phenomena recorded in the lexicon is given, as well as a thorough description of the encoding of all the linguistic information. First we will give a formal description of the format of a frame and then we will explain the meaning of single fields. After that we will describe in depth what kinds of reflexive and reciprocal verbs we distinguish in the Czech language and how we encode them in the lexicon. Then we will deal with diatheses covered by the lexicon and finally we will discuss the so called *equi* and *raising* verbs.

4.1. Format of a lexical entry

A lexical entry contains a lemma and its frame.[1] The term *frame* usually denotes all complementations of a verb in one meaning. The existence of another frame then signals a new meaning. There are, however, variants of surface realizations of functors—in such a case we do not introduce a new meaning but we merge the variants into one frame. In our lexicon, the frame contains all the variants merged together, and in addition it also includes information on possible diatheses. As it is not always possible to accomodate all the combinations of surface realizations and diatheses into one frame, we may be forced to split one meaning into several lexical entries. The identification of one lexical meaning is then provided by indices (different from the indices from the morphological lexicon) attached by ˜. Examples of lexical entries are shown in (16).

```
(16) adresovat   Act[Noun+Nom] Pat[Noun+Acc] \
                    Addr[Noun+Dat|Noun+Acc+na|Noun+Acc+pro] \
                    PeriphPass ReflPass                          (address)
     stát-2~1    Act[Noun+Nom] Gram:{Refl[se]} NoPass            (happen)
     stát-2~2    Act[Noun+Nom] Pat[Noun+Ins] Gram:{Refl[se]} NoPass
                                                                 (become)
     stát-3~1    Act[Noun+Nom] ReflPass                          (stand)
```

[1] As we expect our lexicon to be used together with the morphological lexicon created by J. Hajič (Hajič, 1994) the lemmas must be identical with the lemmas of the morphological lexicon. This means that lemmas must contain the same indices as the morphological lexicon (e.g. *stát-2* (happen), *stát-3* (stand), *stát-4* (cost), etc.). Furthermore, lemmas of reflexive verbs do not contain the reflexive particle (e.g. *stát se* (happen) will have the lemma *stát-2*).

```
stát-3~2    Act[Noun+Nom] Pat[Noun+Acc+o] ReflPass          (long for)
stát-3~3    Act[Noun+Nom] Pat[Noun+Ins+za] Gram:{Refl[si]} NoPass
                                                        (be convicted)
stát-4      Act[Noun+Nom] Pat[Num+Acc] NoPass                   (cost)
učit~1      Act[Noun+Nom] Gram:{Refl[se]} NoPass               (learn)
učit~2      Act[Noun+Nom] Pat[Noun+Acc] Addr[Noun+Acc] NoPass
učit~2      Act[Noun+Nom] Pat[Noun+Dat] Addr[Noun+Acc] \
               PeriphPass ReflPass                             (teach)
```

The verb *adresovat* (address) has only one lemma in the morphological lexicon and only one meaning. The verb *stát* has three different lemmas in the morphological lexicon—one for the reflexive verb *stát se* and two for the non-reflexive verb *stát* (the reflexive verb *stát si* is morphologically covered by the non-reflexive verb *stát*). The reflexive verb is split into two entries with two different meanings in our lexicon (stát-2~1 and stát-2~2), the meanings of the non-reflexive verbs are partly differentiated by the indices from the morphological lexicon, so we have to decide which of the "morphological" meanings will be split. The verbs *učit se* and *učit* have only one entry in the morphological lexicon, but we have to introduce two meanings for them. The second meaning (*učit*—teach) must itself be split into two frames, as the frame variant with two Accusatives does not allow for the formation of a passive, while the variant with Accusative and Dative allows for the formation of both periphrastic and reflexive passive.[2]

The frame is separated from the lemma by a tabulator. A frame has the following format:

<*voice*><*reflexivity*><*subject*>?[<*functor*><*grammatemes*>]*<*diatheses*>

A frame starts with a voice marker, which is obligatory. Then follows a marker for reflexivity, which is also obligatory. The subject marker may be missing, as there exist verbs without a subject. After the subject marker, a list of functors and their corresponding grammatemes follows. This list can be empty, as we suppose that there are verbs with an empty frame (the obvious candidates, meteorological verbs, however, do not belong to this category, as they need an obligatory modification of the location; e.g. *pršet (kde)*—rain). The frame ends with markers of possible diatheses.

In the following sections, single parts of a frame will be described in detail.

4.1.1. Voice

The voice marker shows whether the frame is in active voice or in passive voice. The passive frames are listed only rarely, as normally they are "derived" from the active frames. The marker occupies one position and currently the following characters are used:

[2]In fact, the variant with Patient realized by Accusative also allows passives, but only if the Addressee is general. We will show later how we encode passive which needs special treatment.

`R` — active frame

`P` — irregular passive frame

All frames in example (16) will have the marker R. The missing passives of the verb *učit* (*matematika je učena*, *matematika se učí*—mathematics is taught) will be encoded in a frame starting with P.

4.1.2. Reflexivity

The reflexivity abbreviation marks the type of reflexive particle; reflexive pronouns are treated as a value of the grammateme *semantic features* (see below). The possible values are:

`--` — no reflexive particle

`SE` — reflexive tantum with particle *se* or reflexive passive

`DE` — derived reflexive with particle *se*

`se` — reflexive with optional particle *se*

`SI` — reflexive tantum with particle *si*

`DI` — derived reflexive with particle *si*

`si` — reflexive with optional particle *si*

The term *reflexive with optional particle* denotes verbs that can occur with or without the reflexive particle in the same meaning, and both these possibilities are grammatical. There are two possible sources of such verbs: one source are reflexives tantum with deletable *se*, the other source are non-reflexive verbs with added *se*. As it is diffucult to say which is which, and as we are not interested in 'etymology' of single frames we merged the two source groups into one. It is usually true for such verbs that the reflexive particle is optional for some meanings, and obligatory or impossible for others.

4.1.3. Subject

The subject marker points to the member of the frame which is the subject (if the construction has a subject, otherwise this marker is missing). For an active frame, it points to the subject of an active sentence. When a passive frame is derived from the active one, this pointer changes so that it points to the subject of a passive sentence. In a passive frame, this pointer must point to the subject of a passive sentence.

`s[i1]` — Actor is the subject

`s[a1]` — the subject is raised from Actor's frame

4.1.4. Functor

Functor is a one-character abbreviation of the functor on the tectogrammatical level. All the values are listed in Appendix B. Here we list only abbreviations of inner participants.

`1` — Actor

`2` — Patient

`3` — Addressee

`4` — Origin

`5` — Effect

`0` — no participant; used in frames of raising verbs

4.1.5. Grammatemes

The list of grammatemes determines the morphemes on the morphemic level. There can be several possible surface realizations which are separated by a vertical bar (|). The notation of grammatemes is taken from the source dictionary, but the repertory is enhanced by some features not previously taken into consideration. The grammatemes are given below:

`h` — 'semantic' features; their description is given above in Chapter 3 and in Appendix B. We added the value S for a short reflexive pronoun and we allow grouping of all four nominal values together (hPTSR). More details are discussed below in Section 4.2. Another value which we added is the value Z for pronouns. An explanation why we need this value is given in footnote 9 in Section 4.2. We also added the value G for general participants and E for deleted (empty, erased) deletable modifications in certain secondary frames. Another value which was added is C for the direct speech.

`c` — morphological case; possible values are 1, 2, 3, 4, 6, 7

`r` — preposition; prepositions are enclosed by curly brackets ({na}, {o}, etc.)

The following grammatemes were added:

`n` — number; the values are S and P for singular and plural, respectively. This grammateme was added to the original BRIEF attributes because of the proper treatment of reciprocal verbs (see Section 4.2).

`x` — reciprocal coreference; the value points to a functor which is coindexed with the functor containing this grammateme. It was added because of reciprocal verbs.

`a` — subject raised to object position; the value points to the embeded clause from which the subject was raised

`q` — subject- or object-control

`p` — "patient" control

`t` — "addressee" control

`d` — diatheses of embeded infinitive; the values are identical with values of the "main" frame

`l` — required lexeme

`m` — modality marker

Their meaning will be explained in the further text.

The whole list of grammatemes is closed in brackets whose shape determines whether the participant (functor) is obligatory, general or obligatory and deletable, optional:

`[ ]` — obligatory

`( )` — obligatory inner participant which can be realized as general participant, or obligatory and deletable free modification

`< >` — optional

In FGD, only obligatory free modifications are considered to belong to a verb frame. In practical applications, however, it may be useful to include also free modifications which occur frequently with a given verb. M. Straňáková (Straňáková-Lopatková, 2001) introduced the term *quasi-valency* for such free modifications and we will mark them as optional free modifications. The term *quasi-valency* will be used in one more meaning: it will denote a free modification which only allows some of the surface realizations typical for that free modification.

4.1.6. Diatheses

Many of the diatheses, especially passive constructions, are derived regularly, as will be shown in Chapter 4.3. This is why we do not list all of them in the lexicon but we rather mark single frames with a sign showing which types of diatheses can be derived from the active frame. We adopted special marks for single types of diatheses and we concatanate them to strings.

`%` — periphrastic passive can be derived

(17) a. *Nájemníci*$_{Act}$ *žádají správcovou*$_{Addr}$ *[o přístup na dvůr]*$_{Pat}$.
Tenants$_{Nom}$ ask caretaker$_{Acc}$ for access to yard.

b. *Správcová*$_{Addr}$ *je (nájemníky*$_{Act}$*) žádána [o přístup na dvůr]*$_{Pat}$.
Caretaker$_{Nom}$ is (tenats$_{Ins}$) asked for access to yard.

`$` — reflexive passive is possible

(18) a. *[O tom]*$_{Pat}$ *právě mluvíme.*
About it$_{Loc}$ just now speak$_{1Pl}$.

b. *[O tom]*$_{Pat}$ *se právě mluví.*
About it$_{Loc}$ SE just now speaks.
'It is being spoken about just now.'

`@` — no passive is possible (most reflexives tantum)

(19) a. *Strašidel*$_{Pat}$ *se nebojíme.*
Ghosts$_{Gen}$ SE fear$_{Neg1Pl}$.
'We don't fear ghosts.'

b. * *Strašidel není báno.*
Ghosts$_{Gen}$ is$_{Neg}$ feared.

c. * *Strašidla nejsou bána.*
Ghosts$_{Nom}$ are$_{Neg}$ feared.

d. * *Strašidel se nebojí.*
Ghosts$_{Gen}$ SE fears$_{NegSgNeut}$.

The sentence (19d) is of course grammatical if we understand it as an active sentence with dropped personal pronoun.

`#` — constructions with *mít* (they are discussed in Section 4.3)

(20) a. *Maminka slíbila Péťovi hračku.*
Mummy$_{Nom}$ promised Péťa$_{Dat}$ toy$_{Acc}$.

b. *Péťa má slíbenu hračku.*
Péťa$_{Nom}$ has promised$_{PrtcplFemAcc}$ toy$_{FemAcc}$.

`~` — constructions with *dostat*

(21) a. *Maminka slíbila Péťovi hračku.*
Mummy$_{Nom}$ promised Péťa$_{Dat}$ toy$_{Acc}$.

b. *Péťa dostal slíbenu hračku.*
Péťa$_{Nom}$ got promised$_{PrtcplFemAcc}$ toy$_{FemAcc}$.

c. *Učitelka vynadá neposlušným dětem.*
Teacher$_{Nom}$ scolds disobedient children$_{Dat}$.

d. *Neposlušné děti dostanou vynadáno.*
Disobedient children$_{Nom}$ get$_{Fut}$ scolded.

* — another type of construction with *mít*. Linguists consider this construction to be rather a special verb tense (see Hausenblas, 1963) or they include it in a system of aspects (see Panevová, 1971). We will discuss this in Section 4.3.

(22) a. *Kuchařka uvařila oběd.*
Cook$_{FemNom}$ cooked lunch$_{Acc}$.

b. *Kuchařka má oběd uvařen.*
Cook$_{FemNom}$ has lunch$_{MascAcc}$ cooked$_{PrtcplMascAcc}$.

c. *Kuchařka má uvařeno.*
Cook$_{FemNom}$ has cooked$_{PrtcplNeut}$.

The whole frame then looks as in (23):

(23) a. **`akumulovat`** `R--s[i1]1(hPTc1)2[hTc4]%$` (accumulate st)

b. **`kazit~2`** `RDEs[i1]1[hTc1]@` (decay)

c. **`přihlásit~1`** `R--s[i1]1(hPc1)2[hPTSRc4]A[hTc2r{do}|hTc4r{na}]%$` (enroll sb/st where)

d. **`vyhrát~3`** `R--s[i1]1(hPc1)2[hTc4]4<hPc6r{na}>%$` (win st of sb)

e. **`tázat`** `P--s[i3]1(hPc7)2(sF|sR|hPTc4r{na})3[hPc1]` (ask)

The frame in (23a) is a frame of a transitive verb. The frame has two actants, Actor (1) and Patient (2). Patient is obligatory ([ ]), while Actor can be general (( )). The Actor is realized as a noun (a person or a thing) in Nominative (hPTc1), Patient is realized as a noun (a thing) in Accusative (hTc4). The subject of an active sentence is Actor (s[i1]). Both periphrastic (%) and reflexive ($) passives are possible.

The frame in (23b) is a frame of a derived reflexive (DE). The frame contains only obligatory Actor which is realized as a noun (a thing) in Nominative (1[hTc1]). There is no possibility of passive voice (@).

The frame in (23c) is a frame of transitive verb with quasi-valency. The Patient can be also realized by a reflexive pronoun (both short and long form—hPTSR). The quasi-valency is a free modification with the meaning *where*, but not all realizations of this meaning can be applied. For example preposition *pod* (under) plus a noun in Accusative are unacceptable.

		se	*si*
se (*si*) is a complementation of the verb	true reflexive	*mýt se*	*koupit si jízdenku*
	reciprocal	*milovat se*	*psávat si*
	dative of possesion	Ø	*držet si klobouk*
se (*si*) changes the meaning of the non-reflexive verb	passive	*obilí se mlátí*	Ø
	derived lexical meaning	*větev se zlomila* *vrátit se, učit se*	*zlomit si ruku* *sednout si*
	independent category	*ta kniha se dobře čte* *chce se mi spát*	Ø
se (*si*) is a particle	reflexive tantum	*smát se*	*stěžovat si*

Table 4.1.: Taxonomy of reflexive verbs

In (23d) we can see a frame with obligatory Patient and generalizable Origin.

The frame in (23e) is an example of an irregular passive frame. The generalizable Actor is realized as a noun in Instrumental, Patient as Accusative with the preposition *na* and Addressee as Nominative.

4.2. Reflexivity

In Czech, there is a reflexive pronoun *se* which has several different forms for different cases which can be stressed (long) or unstressed (short). There also exist two reflexive particles which are homonymous with the unstressed reflexive pronoun in Dative (*si*) and Accusative (*se*). In linguistic theory, we distinguish several types of reflexive verbs, but in the lexicon some distinctions will be omitted. We will base our work on taxonomy by K. Králíková (1981) in Table 4.1, but we will adapt it slightly.

In the lexicon, Dative of possesion will not be listed as it does not belong to a verb frame (it is treated as a free modification Beneficiary). The reflexive passive belongs among diatheses and will be treated by the respective rules. The "independent category" will be treated as a diathesis as well.

4.2.1. True reflexive with *se*

True reflexive with *se* is a verbs with reflexive pronoun in Accusative. The pronoun occupies a place of an actant and expresses the coreference of this actant with subject. In most cases it is possible to use the stressed form of the pronoun as well, though the meaning is not fully synonymous.

Some authors doubt about the group of *true reflexive* verbs. It was proposed already by B. Havránek (1928), that *se* in such constructions as *mýt se* (wash self) is not a pronoun (representing a member of a verb frame), but rather a reflexive particle. The group of true reflexive verbs would contain only a couple of constructions like *vidět se v zrcadle* (see oneself in a mirror), *udělat se samostatným* (make oneself independent), etc. This view is supported nowadays by K. Oliva (2000) who shows the behaviour of the particle *se* in opposition with the long form of the pronoun *sebe* and with the short personal pronouns:

(24) a. $\_{}_i$ *Umyl* *se*$_i$ *celý*$_i$.
_ Washed$_{3Sg}$ SE whole$_{Nom}$.

b. $\_{}_i$ *Umyl* *sebe*$_i$ *celého*$_i$.
_ Washed$_{3Sg}$ self$_{Acc}$ whole$_{Acc}$.

c. $\_{}_i$ *Umyl* *ho*$_j$ *celého*$_j$.
_ Washed$_{3Sg}$ him$_{Acc}$ whole$_{Acc}$.

K. Oliva in his work claims that the verb frames with stressed and unstressed forms of the pronoun *se* are actually two different frames. The verb with unstressed form of the pronoun behaves like reflexive tantum and the pronoun is in fact a particle. There is some evidence for this claim, e.g. the stressed and unstressed forms of the pronoun are not freely replaceable, as shown in (25a) and (25b).[3]

(25) a. *Chce se někdo bát sebe?*
Wants SE someone$_{Nom}$ fear$_{Inf}$ self$_{Gen}$?
'Does anyone want to be afraid of himself/herself?'

b. * *Chce se někdo bát se?*
Wants SE someone$_{Nom}$ fear$_{Inf}$ SE?

For us, the important criterion is whether the form *se* (or *si*) can be replaced by the long form *sebe* (*sobě*), and whether the constructions with the short reflexive pronouns are similar to constructions with other (short) pronouns. If we adopted the view that *se* is a particle with no representation on the tectogrammatical level we would get two different descriptions of sentences which we consider nearly synonymous:

(26) a. *umyl_se*
Jan$_{Act}$

Jan se umyl.
Jan SE washed.

b. *umyl*
Jan$_{Act}$ *sebe*$_{Pat,Cor}$

Jan umyl sebe.
Jan washed self.

[3] We can argue that the verb *bát se* requires an object in Genitive, and thus these examples do not prove anything about verbs with Patient in Accusative.

Therefore, we do not go as far as Oliva and still consider the short form to be a pronoun (not a particle), but we are aware of the fact that the short and long forms of the pronoun are not always replaceable and thus, in the lexicon, both possibilities must be explicitly mentioned. We enhanced the repertory of 'semantic' features and added the feature S for the short form of the reflexive pronoun. The frames for the verb *umýt* then will have the following form:

(27) `umýt      R--s[i1]1(hPc1)2[hPTSRc4]3(hPSRc3)%$`[4]

4.2.2. **True reflexive** with ***si***

True reflexive with *si* is a verb with a reflexive pronoun in Dative. The pronoun occupies the place of an actant and expresses the coreference of this actant with the subject. The reflexive pronoun in Dative also has a short and a long form (*si* and *sobě*), which can be used in the same constructions.[5]

(29) a. *Každý*$_{Act,i}$ *si*$_{Addr,i}$ *koupí jízdenku*$_{Pat}$.
Everyone SI buys ticket.

b. `koupit      R--s[i1]1(hPTc1)2[hTc4]3(hPSRc3)%$`

4.2.3. **Reciprocal verbs** with ***se***

Reciprocal verb with *se* is a reflexive verb where the reflexive pronoun has the meaning 'each other'. Similarly to the situation with true reflexives, Actor is identical with other participant (usually Patient) and the reflexive pronoun expresses this. The difference is that there must be at least two bodies participanting in the action and their roles are cross-linked. In fact, there are two actions occuring at the same time, in one of them

[4] The notation in (27) also allows realizations

(28) a. * *umýt* *si* *se*
wash self$_{Dat}$ self$_{Acc}$

b. ? *umýt* *si* *sebe*
wash self$_{Dat}$ self$_{Acc}$

c. ? *umýt* *se* *sobě*
wash self$_{Acc}$ self$_{Dat}$

d. ? *umýt* *sobě* *sebe*
wash self$_{Dat}$ self$_{Acc}$

which can be handled by a general rule of grammar saying that two (short) reflexive pronouns cannot occur as realizations inside one verb frame.

[5] The two forms of the pronoun have of course different roles in topic-focus articulation and so we cannot say that they are synonymous, but in this work we are mainly interested in *grammaticality* of verb constructions and we do not try to describe all semantic nuances of proposed frames.

	koulovat (se) (snowball)	*hašteřit se* (wrangle)	*soutěžit* (compete)
	A kouluje B B kouluje A		
reciprocal	AB se koulují	AB se hašteří	AB soutěží
reciprocal	A a B se koulují	A a B se hašteří	A a B soutěží
reciprocal	A s B se koulují	A s B se hašteří	A s B soutěží
reciprocal	A se kouluje s B	A se hašteří s B	A soutěží s B

Table 4.2.: Three types of reciprocal verbs

the participant i is Actor and participant j is Patient and in the other action the roles are exchanged.

When we examine so called reciprocal verbs closer we discover that there are three types of them. The first type (represented by the verb *koulovat (se)*) was described in the previous paragraph. The second type is reflexive tantum with reciprocal meaning (inherently reciprocal verb). The reciprocal meaning is manifested by obligatory participant with the surface form *s kým* (with whom).[6] The third type is a "plain" verb with reciprocal meaning. The three types are shown in Table 4.2.

All these types were described by J. Panevová(1999), with a proposal how to encode the information in a lexicon. Her work, however does not suggest structures for sentences with reciprocal verbs. We try to make a proposal of the structures and we will compare them to structures proposed in (Hajičová et al., 2000). Our proposal is shown in (30):

(30) a. *koulovat*

chlapci$_{Act,Pat,RECP}$ se$_{Act,Pat,RECP}$

Chlapci se koulují.
Boys$_{Nom}$ SE snowball.

b. *koulovat*

COORD se$_{Act,Pat,RECP}$

Petr$_{Act,Pat,RECP}$ *Pavel*$_{Act,Pat,RECP}$

Petr a Pavel se koulují.
Petr$_{Nom}$ and Pavel$_{Nom}$ SE koulují.

Petr s Pavlem se koulují.
Petr$_{Nom}$ with Pavel$_{Ins}$ SE snowball.

c. *koulovat_se*

Petr$_{Act,RECP}$ *Pavel*$_{Pat,RECP}$

Petr se kouluje s Pavlem.
Petr$_{Nom}$ SE snowballs with Pavel$_{Ins}$.

[6] This means that the answer for the question *S kým to dělal?* (With whom did he do that?) cannot be *S nikým. Sám.* (With no-one. Alone.)

d. *koulovat_se*

$Petr_{Act,RECP}$ $\quad$ $\mathrm{GNRL}_{Pat,RECP}$

? *Petr se kouluje* $\_{GNRL}$.
Petr_{Nom} SE snowballs.

It may be surprising that *se* is treated as a pronoun in (30a) and (30b), and as a particle in (30c) and (30d). This is a result of applying the criteria which we used already for the true reflexives:

(31) a. *Chlapci koulují sebe (navzájem).*
Boys snowball self (each other).
'Boys snowball each other.'

b. *Petr a Pavel koulují sebe (navzájem).*
Petr and Pavel snowball self (each other).

c. *Petr s Pavlem koulují sebe (navzájem).*
Petr with Pavel snowball self (each other).

d. **Petr kouluje sebe s Pavlem.*
Petr snowballs self with Pavel.

In (31a)–(31c), the short form of the pronoun *se* can be replaced by the long form (which indicates that it is really a pronoun), while in (31d) this replacement is not possible (which indicates that *se* is a particle).

Our proposal differs from solution used in a manual for tectogrammatical tagging of the Prague Dependency Treebank (Hajičová et al., 2000). There the authors recommend to tag the above sentences as shown in (32).

(32) a. $Chlapci_{Act}$ $se_{Recp,Pat}$ *koulují.*
Boys SE snowball.

b. $Petr_{Act}$ *a* $Pavel_{Act}$ $se_{Recp,Pat}$ *koulují.*
Petr and Pavel SE snowball.

c. $Petr_{Act}$ *s* $Pavlem_{Pat}$ *koulují_se.*
Petr with Pavel snowball.

d. $Petr_{Act}$ *kouluje_se s* $Pavlem_{Pat}$.
Petr snowballs with Pavel.

The two solutions differ in assignment of functors to single participants of a sentence, and in understanding of coordination. In PDT, it is also supposed that the relation of reciprocity always includes Actor, and so it gets no Recp marker. But in the end, the proper treatment of these construction on the tectogrammatical level is not the main point for us. Our task is to encode a lexical entry in such a way that the proper treatment is enabled. This means that we have to mark the reciprocity of the two participants. We decided to introduce a new grammateme `x` whose value is the coreferential functor. In example (33b), there is Actor in plural,[7] and Patient is realized by the reflexive pronoun *se* (both short and long form). It is marked as reciprocally coreferential with Actor. In example (33c), Patient has morphological realization by Instrumental+*s*, and it is also reciprocally coreferential with Actor.

(33) a. `R--s[i1]1[hPc1]2[hPc4]%$`

b. `R--s[i1]1[hPc1nP]2[hSRc4x1]@`

c. `RDE1[hPc1]2(hPRc7{s}x1)@`

The frame in (33b) corresponds to the sentences (30a) and (30b). The frame in (33c) corresponds to the sentences (30c) and (30d).

For the sake of completeness we also show the frame of inherently reciprocal verbs *hašteřit se* and *soutěžit*:

(34) a. *Chlapci*$_{ActPatRecp}$ *se hašteří.*
Boys$_{Nom}$ SE wrangle.

b. *Petr*$_{ActPatRecp}$ *a Pavel*$_{ActPatRecp}$ *se hašteří.*
Petr$_{Nom}$ and Pavel$_{Nom}$ SE wrangle.

c. *Petr*$_{ActPatRecp}$ *s Pavlem*$_{ActPatRecp}$ *se hašteří.*
Petr$_{Nom}$ with Pavel$_{Ins}$ SE wrangle.

d. **hašteřit** `RSEs[i1]1[hPc1nP]2[x1]@`

e. *Petr*$_{ActRecp}$ *se hašteří s Pavlem*$_{PatRecp}$.
Petr$_{Nom}$ SE wrangles with Pavel$_{Ins}$.

f. *Chlapci*$_{ActRecp}$ *se hašteří* $\_{}_{GnrlPatRecp}$.
Boys$_{Nom}$ SE wrangle.

g. **hašteřit** `RSEs[i1]1[hPc1]2(hPc7{s}x1)@`

(35) a. *Chlapci*$_{ActPatRecp}$ *soutěží.*
Boys$_{Nom}$ compete.

b. *Petr*$_{ActPatRecp}$ *a Pavel*$_{ActPatRecp}$ *soutěží.*
Petr$_{Nom}$ and Pavel$_{Nom}$ compete.

[7] The plural here means *semantic* plural, not grammatical. It can be realized as a noun in plural, or as a coordination or as a noun with meaning of a group, e.g. *třída* (class).

	povídat (si) (chat, *imperf.*)	*popovídat si* (chat, *perf.*)
	A povídá B o ... B povídá A o ...	
reciprocal	AB si povídají o ...	AB si popovídají o ...
reciprocal	A a B si povídají o ...	A a B si popovídají o ...
reciprocal	A s B si povídají o ...	A s B si popovídají o ...
reciprocal	A si povídá s B o ...	A si popovídá s B o ...

Table 4.3.: Reciprocal verbs with *si*

c. *Petr*$_{ActPatRecp}$ *s Pavlem*$_{ActPatRecp}$ *soutěží.*
Petr$_{Nom}$ with Pavel$_{Ins}$ compete.

d. `soutěžit` `R--s[i1]1[hPc1nP]2[x1]@`

e. *Petr*$_{ActRecp}$ *soutěží* *s Pavlem*$_{PatRecp}$.
Petr$_{Nom}$ competes with Pavel$_{Ins}$.

f. *Chlapci*$_{ActRecp}$ *soutěží* $\_{GnrlPatRecp}$.
Boys$_{Nom}$ compete.

g. `soutěžit` `R--s[i1]1(hPc1)2(hPc7{s}x1)$`

From the above description it follows that there is no need to introduce a new mark for “reciprocal” *se* as it is possible to use other defined markers.

4.2.4. Reciprocal verbs with *si*

Reciprocal verb with *si* is a reflexive verb where the reflexive pronoun has the meaning ‘(with/to) each other’. The properties are similar to properties of reciprocal verbs with *se*; the difference is that the functors assigned to the participants of the action are Actor and Addressee. The types of reciprocal verbs with *si* are shown in Table 4.3.

(36) a. `R--1(hPc1)2[hTc6r{o}]3(hPc3)$` (*povídat*)

b. `R--1[hPc1nP]2[hTc6r{o}]3[hSRx1]@` (*povídat si*, *popovídat si*)

c. `RDI1[hPc1]2[hTc6r{o}]3(hPc7{s}x1)@` (*povídat si*)

(37) a. `RSI1[hPc1nP]2[hTc6r{o}]3[x1]@` (*popovídat si*)

b. `RSI1[hPc1]2[hTc6r{o}]3(hPc7{s}x1)@` (*popovídat si*)

4.2.5. **Reflexive tantum** with *se*

Reflexive tantum with *se* is a verb which has an obligatory reflexive particle *se*. This particle has no representation on the tectogrammatical level.

(38) a. *Helena*$_{Act}$ *se směje všemu*$_{Pat}$.
Helena$_{Nom}$ SE laughs everything$_{Dat}$.
'Helena laughs at everything.

b. *František*$_{Act}$ *se nebojí ničeho*$_{Pat}$
František$_{Nom}$ SE fears nothing$_{Gen}$.
'František is not afraid of anything.'

Frames of verbs from above examples will look as follows:[8]

(39) a. `smát` `RSE1[hPc1]2<hPTRc3>@`

b. `bát` `RSE1[hPc1]2(hPTRc2|sD|sF|sU)@`

4.2.6. **Derived reflexive verbs** with *se*

This category contains verbs which behave like reflexive tantum but they have origin in true reflexive verbs. Their lexical meaning, however, changed so that the cannot be understood as true reflexives any more. For example the verb *rozčílit se* (get angry) could be understood as true reflexive, as it is possible to say *rozčílit sám sebe* (make angry oneself), but the meaning is different (as the translation also shows). Beside it, the verb *rozčílit se* has only Actor in its frame, while *rozčílit koho/sebe* has Actor, Patient and Addressee. The verb *rozčílit* then will have two meanings with two frames, as shown in example (40).

(40) a. `rozčílit~1` `R--1[hPTc1]2(hTc7)3[hPTRc4]@`

b. `rozčílit~2` `RDE1[hPc1]@`

4.2.7. **Reflexive tantum** with *si*

Reflexive tantum with *si* is a verb which has an obligatory reflexive particle *si*.

(41) a. *Nájemníci*$_{Act}$ *si stěžují [na správcovou]*$_{Pat}$
Tenants$_{Nom}$ SI complain about caretaker.

b. `stěžovat` `RSI[hPc1]2[hPTRc4r{na}]@`

[8] In the frame of the verb *bát se* the realization by infinitive is missing. This is because the infinitive needs special treatment—raising or control must be marked. This will be discussed in Section 4.4 and thus we did not want to obscure this example.

4.2.8. **Derived reflexive verbs** with ***si***

This category is similar to derived reflexive verbs with *se*.

(42) a. *Děti*$_{Act}$ *si hrají [na indiány]*$_{Pat}$
Children$_{Nom}$ SI play at indians$_{Acc}$.

b. `hrát       RSI[hPc1]2<hPTRc4r{na}>@`

4.2.9. **Reflexive** with optional ***se***

This is a verb with reflexive particle *se* which is not obligatory. It is usually true for such verbs that the reflexive particle is optional for some meanings, and obligatory or impossible for others.

(43) a. *Na co*$_{Pat}$ *(se) koukáš?*
On what$_{Acc}$ (SE) look$_{2Sg}$?
'What are you watching?'

b. `koukat~1    Rse1[hPc1]2(hPTRc4r{na})$`

c. *Kouká ti*$_{Ben}$ *podolek*$_{Act}$.
Looks you$_{Dat}$ shirt-tail$_{Nom}$.
'Your shirt-tail is showing.'

d. * *Kouká se ti podolek.*
Looks SE you$_{Dat}$ shirt-tail$_{Nom}$.

e. `koukat~2    R--1[hTc1]@`

Reciprocal verb with optional ***se***

Some of the the reflexive verbs with optional *se* can also be inherently reciprocal.

(44) a. *Vy už (se) spolu nekamarádíte?*
You$_{2PlNom}$ already (SE) together hobnob$_{Neg}$?

b. *Já (se) s Jirkou kamarádím!*
I$_{Nom}$ (SE) with Jirka$_{Ins}$ hobnob.

c. `kamarádit    Rse1[hPc1nP]2[x1]@`

d. `kamarádit    Rse1(hPc1)2[hPc7{s}x1]@`

4.2.10. **Reflexive** with optional ***si***

Reflexive with optional*si* is a verb with reflexive particle*si* which is not obligatory.

(45) a. *Aleš*$_{Act}$ *(si) myslí, [že Jiřina nepřijde]*$_{Pat}$.
Aleš$_{Nom}$ (SI) thinks that Jiřina comes$_{FutNeg}$.

b. *Aleš*$_{Act}$ *si to*$_{Pat}$ *nemyslí.*
Aleš$_{Nom}$ SI it$_{Acc}$ thinks$_{Neg}$.

c. *Co*$_{Pat}$ *(si)* myslí *Aleš*$_{Act}$*?*
What$_{Acc}$ (SI) thinks Aleš$_{Nom}$?

In example (45) we can see that the verb *myslet si* does not require the particle obligatorily if it is complemented by a clause. It requires the particle, however, if the complementation is realized by a pronoun.[9]

On the other hand, the particle *si* cannot occur if we use the verb in its intransitive meaning or in the meaning 'have in mind'.

(46) a. *Myslím, tedy jsem.*
Think$_{1Sg}$, then am.
'Cogito, ergo sum.'

b. * *Myslím si, tedy jsem.*
Think$_{1Sg}$ SI, then am.

c. *Co*$_{Pat}$ *tím myslíš?*
What$_{Acc}$ it$_{Ins}$ think$_{2Sg}$?
'What do you mean by it?'

d. * *Co*$_{Pat}$ *si tím myslíš?*
What$_{Acc}$ SI it$_{Ins}$ think$_{2Sg}$?

The verb *myslet (si)* then will need several frames which will express the behaviour of the particle *si*.

(47) a. `myslet~1    Rsi1[hPc1]2[sD]@`

[9]Here we can observe another interesting phenomenon. A complementation realized by a noun in a certain case includes also realizations by pronouns in that case. On the other hand, the pronouns can also stand for clauses and infinitives, even in such frames where no noun can occur. This was the reason why we introduced a new 'semantic' feature Z for such pronouns. The set of possible realizations covers interrogative pronoun *co* (what), demonstrative pronoun *to* (that), totalizer *všechno* (everything), indefinite pronoun *leccos* (all sorts of things), etc. Not all these pronouns are applicable in all frames as alternatives to clauses/infinitives, but a detailed research is beyond the scope of this work. We will use the single value Z and we can refine our description later.

b. `myslet~1` `RSI1[hPc1]2[hZc4]@`

c. `myslet~2` `R--1(hPc1)$`

d. `myslet~3` `R--1(hPc1)2[hZc4|sD]I(hTc7)&`

e. `myslet~4` `R--1(hPc1)2[hPTc4r{na}]&`

4.2.11. Reflexive passive

Reflexive passive is a construction with the particle *se*. It is one of the possible passive constructions in Czech. This construction is usually derived from the basic active frame and the passive frames are not listed in the lexicon separately.

(48) *Brána se zavírá v devět hodin.*
Gate SE closes at nine o'clock.

This construction will be discussed in detail in Section 4.3.3.

4.2.12. Mediopassive

Mediopassive constructions are a sort of reflexive passive and they will be described later in Section 4.3.

(49) *Z této látky se šije dobře.*
From this fabric$_{Gen}$ SE sews well.
'This is good fabric for sewing.'

In our lexicon these constructions will be treated as reflexive passives. The discussion about this type of construction follows in Section 4.3.3.

4.2.13. Homonymy of reflexive verbs

Some reflexive verbs have several meanings, in which they appear as true reflexive, reciprocal verb, derived reflexive verb, reflexive tantum, or reflexive passive:

(50) a. *Jarda se tím dobře baví.* (true reflexive)
Jarda SE by it well amuses.

b. *S nimi se nebavte!* (reciprocal reflexive)
With them SE do not amuse!
'Don't talk to them!'

c. *Jan a Marie se milují.*
Jan and Marie SE love.
'Jan and Marie love each other.' (reciprocal reflexive) or
'Jan and Marie make love.' (inherent reciprocal, reflexive tantum) or
'Jan and Marie love themselves.' (true reflexive)

d. *Selátko se opéká na rožni.* (reflexive passive or derived reflexive verb)
Piglet SE roasts on spit.

e. *Turisté se opékají na pláži.* (derived reflexive)
Tourists SE roast on beach.

Sometimes, it is difficult for the lexicon maintainer to decide whether a verb is reflexive tantum, true reflexive or a derived reflexive. Here are some hints that can help:

- Reflexive tantum is usually isolated in the lexicon. The verb without the reflexive particle does not exist in the language at all:

  (51) * *smát koho/co* (laugh sb/st)

- The verb with and without the reflexive particle has the same meaning. Then it is either reflexive with optional particle or a reciprocal verb with optional particle.

- The verb requires an answer for the question 'Who with whom?'; then it is a reciprocal verb.

- Actor and the participant in question can be both persons and both short and long forms of the reflexive pronoun can be used as the Patient (Addressee). Then we have true reflexive. We can also use the construction with *sám sebe/sobě*:

  (52) a. *Honza*$_{Act}$ *myje sám sebe*$_{Pat}$.
  Honza$_{Nom}$ washes self$_{Nom}$ self$_{Acc}$.

  b. *Honza*$_{Act}$ *koupí jízdenku*$_{Pat}$ *sám sobě*$_{Addr}$.
  Honza$_{Nom}$ buys ticket$_{Acc}$ self$_{Nom}$ self$_{Dat}$.

- Another test for true reflexive is whether we can form a periphrastic passive with the reflexive pronoun as Actor:

  (53) *Jsem myt (sám) sebou*$_{Act}$.
  Am washed (self$_{Nom}$) self$_{Ins}$.
  'I am washed by myself.'

- If the two above tests fail we may have found derived reflexive :

  (54) a. *Pasáci ženou krávy z pastvy.*
  Herdsmen$_{Nom}$ drive cows$_{Acc}$ from pasture.

b. *Krávy se ženou z pastvy.*
Cows SE drive from pasture.
'Cows are being driven from pasture.' (reflexive passive) or
'Cows rush from pasture.' (derived reflexive)

c. *Pasáci se ženou do hospody.*
Herdsmen SE drive to pub.
'Herdsmen rush into a pub.' (derived reflexive)

- Actor is a person, the participant in question is a thing, subject of the reflexive construction is a thing—then we have derived reflexive or reflexive passive:

(55) a. *Petr*$_{Act}$ *zavírá oči*$_{Pat}$.
Petr$_{Nom}$ closes eyes$_{Acc}$.

b. *Oči*$_{Act}$ *se mu*$_{Pat}$ *zavírají.* (derived reflexive)
Eyes$_{Nom}$ SE him$_{Dat}$ close$_{3Pl}$.

c. *Petr*$_{Act}$ *zavírá bránu*$_{Pat}$.
Petr$_{Nom}$ closes gate$_{Acc}$.

d. *Brána*$_{Act}$ *se (sama) zavírá.* (derived reflexive)
Gate$_{NomPl}$ SE (self) close$_{3Pl}$.

e. *Brána*$_{Act}$ *se zavírá večer.* (reflexive passive)
Gate$_{NomPl}$ SE close$_{3Pl}$ in evening.

f. *Vltava se vlévá do Labe.* (derived reflexive)
Vltava$_{Nom}$ SE flows into Labe$_{Gen}$.
'Vltava joins Labe.'

Derived reflexive verbs or reflexive passives can be used jokingly as true reflexive:

(56) *Náš starodávný dědeček se každé ráno vlévá do vany*
Our ancient grandfather SE every morning flows to bathtub
se studenou vodou ...
with cold water ...
(*J. Vodňanský*)

This is, however, a question of language performance rather than of the lexicon and so we leave this aside.

4.3. Diatheses

Another lexical information useful for language processing is the information about diatheses. The most important diatheses are passive constructions. In Czech there exist two syntactic constructions with passive meaning: the periphrastic passive formed by an auxiliary verb *být* (be) and passive participle, and reflexive passive formed by indicative and the reflexive particle *se*. As both these passives are derived regularly from the active constructions, we will only list the information of what *type* of passive is acceptable for a certain verb and its frame, and we will not list all the passive constructions in our lexicon. Of course, there are exceptions—passive constructions which are derived by exceptional rules—such passives must be listed explicitly (but there will be only single cases of such passive constructions).

Beside periphrastic and reflexive passive, there exist also other types of diatheses which we consider regular. For example, constructions with support verbs *dostat* (get) and *mít* (have) are very frequent. The possibility of marking these types of diatheses in the lexicon will also be discussed.

In our lexicon, we only consider such derived constructions in which the surface syntactic structure is changed. Such constructions as

(57) a. *Bolest probudila Pavla.*
Pain$_{Nom}$ woke Pavel$_{Acc}$.

b. *Marie probudila Pavla.*
Marie$_{Nom}$ woke Pavel$_{Acc}$.

differ in the semantics of subjects. In (57a), the subject has the role of Causer (according to (Daneš et al., 1987a; Štícha, 1984; Grepl and Karlík, 1998)), while in (57b), the subject is Agent. In the FGD approach, however, both subjects have the role of Actor. Both the constructions are identical on the surface level and they only differ in the lexical setting of the subject. F. Štícha in his work argues that constructions with different kinds of Actor (namely Mediator, Agent, Initiator and Causer) trigger different sorts of diatheses, which is shown in (58)–(61).[10,11]

(58) a. *Voda$_{Mediator}$ naplnila jámu.*
Water$_{Nom}$ filled pit$_{Acc}$.

b. *Jáma se naplnila vodou$_{Mediator}$.*
Pit$_{Nom}$ SE filled water$_{Ins}$.
'Pit filled with water.'

c. **Jáma byla naplněna (vodou$_{Mediator}$).*
Pit$_{Nom}$ was filled (water$_{Ins}$).
'Pit was filled by water.'

d. *Jáma byla naplněna (vodou$_{Means}$).*
Pit$_{Nom}$ was filled (water$_{Ins}$).
'Pit was filled with water.'

[10] There is probably a mistake, in the sentence (59c). Following other examples we would rather expect the sentence *Košík byl naplněn Evou.* but this does not affect the correctness of the example.

[11] Sentences in (58)-(61) (a)-(c) are taken from (Štícha, 1984), sentences in (d) were added by us.

(59) a. *Eva$_{Agent}$ naplnila košík prádlem$_{Means}$.*
Eva$_{Nom}$ filled basket$_{Acc}$ linen$_{Ins}$.
'Eva filled the basket with linen.'

b. **Košík se naplnil Evou$_{Agent}$.*
Basket$_{Nom}$ SE filled Eva$_{Ins}$.
'Basket filled with Eva.'

c. *Košík byl naplněn (prádlem$_{Means}$).*
Basket$_{Nom}$ was filled (linen$_{Ins}$).
'Basket was filled with linen.'

d. *Košík byl naplněn (Evou$_{Agent}$).*
Basket$_{Nom}$ was filled (Eva$_{Ins}$).
'Basket was filled by Eva.'

(60) a. *Řečník$_{Initiator}$ naplnil sál vzrušením.*
Speaker$_{Nom}$ filled hall$_{Acc}$ excitement$_{Ins}$.
'The speaker filled the hall with excitement.'

b. **Sál se naplnil řečníkem$_{Initiator}$.*
Hall$_{Nom}$ SE filled speaker$_{Ins}$.
'The hall filled with the speaker.'

c. **Sál byl naplněn řečníkem$_{Initiator}$.*
Hall$_{Nom}$ was filled speaker$_{Ins}$.
'The hall was filled with/by the speaker.'

(61) a. *Obraz$_{Causer}$ naplnil Karla nadšením.*
Picture$_{Nom}$ filled Karel$_{Acc}$ ecstasy$_{Ins}$.
'The picture filled Karel with ecstasy.'

b. **Karel se naplnil obrazem$_{Causer}$.*
Karel$_{Nom}$ SE filled picture$_{Ins}$.
'Karel filled with the picture.'

c. **Karel byl naplněn obrazem$_{Causer}$.*
Karel$_{Nom}$ was filled picture$_{Ins}$.
'Karel was filled with/by the picture.'

We can argue that in (60) and (61) the verb alone does not form the whole lexical unit, but it is rather a part of collocations *naplnit vzrušením* and *naplnit nadšením*. Sentences in (62) show it clearly.

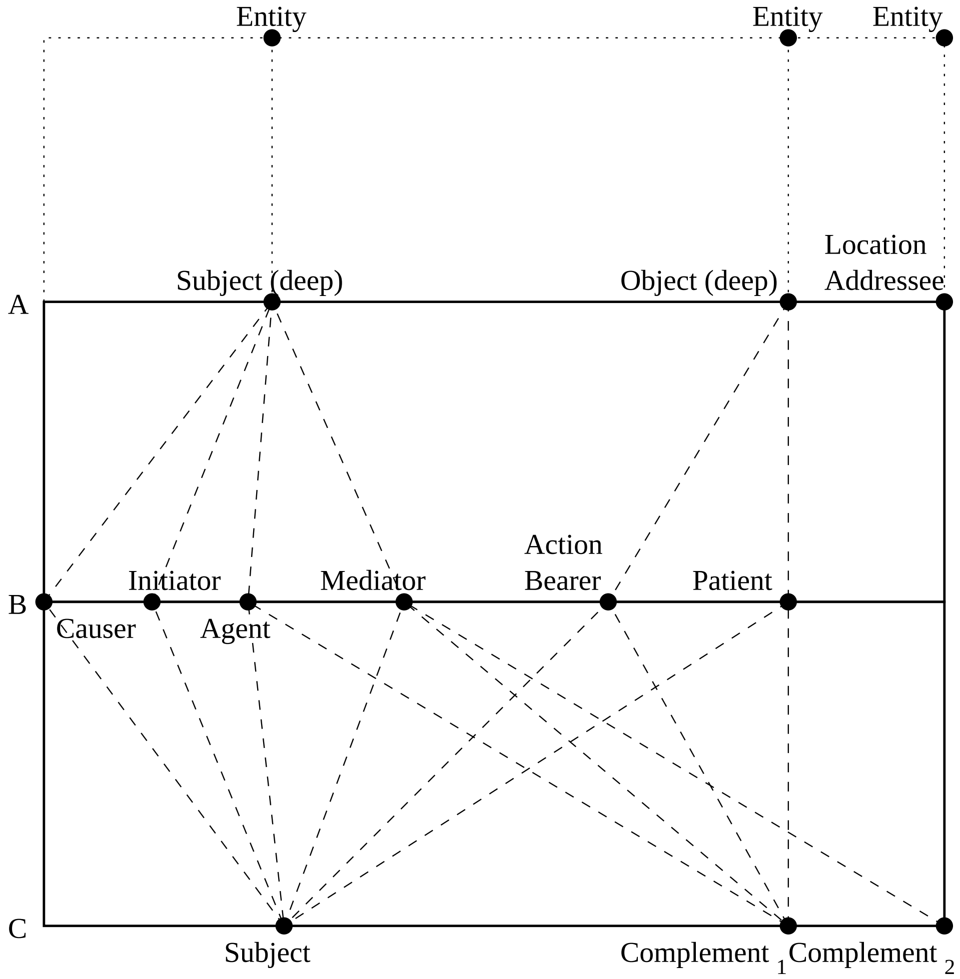


Figure 4.1.: Three-level system

(62) a. *Řečník naplnil sál.*
Speaker filled hall.
'The speaker sold out the hall.'—different meaning of the verb

b. **Obraz naplnil Karla.*
Picture filled Karel.

c. ?*Sál byl řečníkem naplněn vzrušením.*
Hall_{Nom} was speaker_{Ins} filled excitement_{Ins}.
'The hall was filled with excitement by the speaker.'

d. ?*Karel byl obrazem naplněn nadšením.*
Karel_{Nom} was picture_{Ins} filled ecstasy_{Ins}.
'Karel was filled with ecstasy by the picture.'

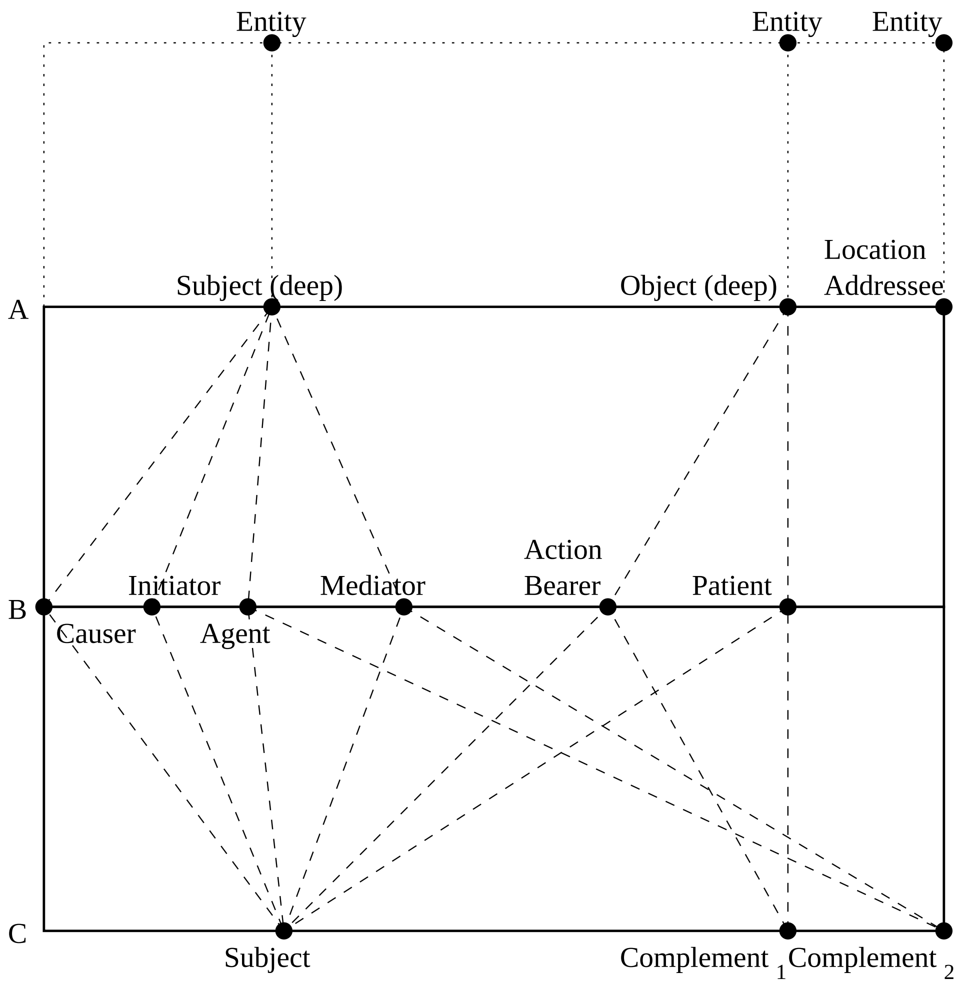


Figure 4.2.: Three-level system revisited

In Štícha's approach, the roles are assigned dynamically to every verb in every sentence, and they are not considered a static property of the verb. The role ascribed to the (surface) subject has an influence on possible diatheses, as was shown above.

Štícha works with a three-level construct of language representation, which is shown in Fig. 4.1. The levels A and B are levels describing the meaning of a sentence, while level C is the level of surface structure. The entities on level A could be compared with functors in FGD. His approach differs from FGD, however, in the treatment of single-valency verbs. In concordance with E. Pauliny (1943) such a verb is understood as occupying both positions of Deep Subject and Deep Object with the single actant.

level B is a level of semantic description. The roles corresponding to Deep Subject are ordered from left to right according to the level of integration of Deep Subject into the situation described in the sentence. Causer is the least integrated, Mediator is the most integrated. Deep Object has two representation on level B: Action Bearer and

Patient. Action Bearer occurs in sentences with Causer, Initiator and Mediator, Patient co-occurs with Agent.

Level C is the level of the surface structures. Štícha does not use the traditional terms as object or adjunct (free modification) but he introduces new terms to avoid confusion with other theories. The term subject is understood as a noun in Nominative (or numerative), the term Complement of the first grade (marked as Complement$_1$) belongs to Accusative (direct object) and Complement of the second grade (Complement$_2$) is characterized by other cases (Dative, Locative or Instrumental).

In Figure 4.1 we can see the possible diatheses. The dashed lines connect corresponding entities; we can see that Deep Object can be realized as surface subject or direct object (but not as indirect object). Deep Subject can be realized as indirect object, only if it has the role of Mediator—this is a little surprising as we would expect that Agent could be realized as Complement$_2$ as well, as in (59d). The realizations of Agent and Mediator as Complement$_1$ (direct object) also seem impossible. The corrected version of the schema is in Figure 4.2.[12]

Beside this, the schema does not say anything about verb forms in the diatheses. It should be said when the verb is in active voice and when in passive. And in the case of passive voice, the form of passive should be marked.

Another classification of diatheses was proposed by P. Karlík (see Grepl and Karlík, 1998). He discriminates *subject diatheses* and *object diatheses*. What Karlík calls object diathesis is understood as a separate frame, in FGD (as shown in (63)), so we will not deal with them.

(63) a. *Petr*$_{Act}$ *napustil vanu*$_{Pat}$ *vodou*$_{Means}$.
Petr poured bathtub water$_{Ins}$.
'Petr filled the bathtub with water.

b. *Petr*$_{Act}$ *napustil vodu*$_{Pat}$ *do vany.*$_{Dir.where}$
Petr poured water to bathtub.

Subject diatheses, in Karlík's concept, differ according to whether the subject of an active sentence has the role of Agent or Causer. The diatheses are shown in Table 4.4.[13]

Examples of deagentive diatheses follow:

[12] We also showed that Causer and Initiator can be realized as Complement$_2$ (see example (62c) and (62d)). The schema then becomes rather trivial: any Deep Subject can be realized as subject or Complement$_2$.

[13] We omitted "active" diatheses, where the structure od the active sentence is not changed, only Agent becomes general:

(64) a. *Zabili nám Ferdinanda.*
Killed$_{3Pl}$ us$_{Dat}$ Ferdinand$_{Acc}$.

b. *S poctivostí nejdál dojdeš.*
With honesty go$_{2Sg}$ farthest.
'Honesty is the best policy.'

	active voice	subject: agent	subject: causer
subject — patient	periphrastic passive	with Instrumental (with Genitive+*od*)	with Instrumental (with Genitive+*od*)
	reflexive passive	+	
	derived reflexive		(with Instrumental)
subject — recipient	support verb *mít*	with Genitive+*od*	with Instrumental (with Genitive+*od*)
	support verb *dostat*	with Genitive+*od*	

Table 4.4.: Subject diatheses

(65) a. *Pokoj byl (uklizečkou*$_{Agent}$*) uklizen.*
Room was (cleaning woman$_{Ins}$) cleaned.
'The room was cleaned by a cleaning woman.'

b. *Pokoj se uklidí.*
Room SE clean$_{Fut3Sg}$.
'The room will be cleaned.'

c. *Žadatel má (od úřadu*$_{Agent}$*) přidělen byt.*
Applicant has (from municipality) assigned$_{PrtcplMascAcc}$ flat$_{MascAcc}$.
'The applicant is assigned a flat by municipality.'

d. *Žadatel dostal (od úřadu*$_{Agent}$*) přidělen byt.*
Applicant got (from municipality) assigned$_{PrtcplMascAcc}$ flat$_{MascAcc}$.
'The applicant was assigned a flat by municipality.'

Examples of decausative diatheses follow:

(66) a. *Pavel byl probuzen (bolestí*$_{Causer}$*).*
Pavel was waken up (pain$_{Ins}$).
'Pavel was waken up by pain.'

b. *Pavel se probudil (bolestí*$_{Causer}$*).*
Pavel SE woke up (pain$_{Ins}$).
'Pavel woke up because of pain.'

c. *Petr má obličej pokryt prachem*$_{Causer}$.
Petr has face covered dust$_{Ins}$.
'Petr's face is covered with dust.'

c. *Před a nepíšeme čárku.*
Before and write$_{Neg1Pl}$ comma.

	active voice	subject: agent	subject: causer
subject — patient	periphrastic passive	with Instrumental (with Genitive+*od*)	with Instrumental (with Genitive+*od*)
	reflexive passive	+	
subject — recipient	support verb *mít*	with Genitive+*od*	
	support verb *dostat*	with Genitive+*od*	

Table 4.5.: Subject diatheses revisited

In FGD, the above examples are parsed differently: *bolest* (pain) in (66b) is understood as Cause (free modification) and the construction in (66c) is understood as derived from another frame (*pokrýt co čím*—cover st with st). If we rewrite the table accordingly (see Table 4.5) we can see that the right column (with causer) can be included in the column of agent.

The purpose of previous discussion was to show that we do not lose any important information if we use the functor Actor for all sorts of the deep subject.

4.3.1. Diatheses encoded in the lexicon

In our lexicon we will consider only such diatheses where functors on the tectogrammatical level are not changed, the sentence only gets a different surface realization. This rules out the so called object diatheses, because they include the change of functors.[14] The so called active diatheses are also ruled out: in Karlík's approach the agent becomes general, but we will rather follow J. Panevová (1984) and we will not consider active sentences as sentences with general Actor. This means that the surface realization is not changed. This also implies that Actor can be general only in passive sentences.

Another type of diatheses which is left aside is the construction of the verb *být* (be) and infinitive:

(67) a. *Odtud je vidět Sněžka / Sněžku / na Sněžku.*
From here is see$_{Inf}$ Sněžka$_{Nom}$/ Sněžka$_{Acc}$/ on Sněžka.
'Sněžka can be seen from here.'

b. *Je znát, že se ochladilo.*
Is know$_{Inf}$ that it got colder.
'One can feel that it got colder.'

In the further text we will deal with the following types of derived constructions:

[14] We can even take this as a proof that in this case the change of surface realization is not a diathesis, but that we encountered a different frame.

- periphrastic passive
- reflexive passive
- mediopassive
- constructions with the verbs *mít* and *dostat*

We will discuss the conditions at which the single types of derived constructions can be formed, and the lexical rules and constraints that can be employed for their construction.

4.3.2. Periphrastic passive

The verb is in the form of the periphrastic passive, the predicate agrees with subject in person, gender and number:

(68) a. *Petr*$_{Act}$ *čte* *knihu*$_{Pat}$.
Petr$_{Nom}$ reads book$_{Acc}$.

b. *Kniha*$_{Pat}$ *je čtena.*
Book$_{FemNom}$ is read$_{PrtcplFemSg}$.

This construction is usually formed from transitive verbs (i.e. verbs with object in Accusative), but there are exceptions. Not all transitive verbs can be passivized (e.g. *mít* 'to have', *dostat* 'to get', etc.), and on the other hand, some verbs without an Accusative object can form passive:

(69) a. *Úřad*$_{Act}$ *vyhověl jeho žádosti*$_{Pat}$. .
Office$_{Nom}$ granted his application$_{Dat}$.
'The office granted his application.'

b. *Jeho žádosti*$_{Pat}$ *bylo (úřadem*$_{Act}$*) vyhověno.*
His application$_{FemDat}$ was (by office$_{Ins}$) granted$_{PrtcplSgNeut}$.
'His application was granted (by the office).'

The subject slot of the passive construction is either filled by the original Accusative object (typically Patient), or it is empty (if the active construction did not contain any Accusative). In the case when the subject is empty or it is a clause (finite or non-finite) the verb shows agreement with neuter singular.

The original subject (Actor) changes its case to Instrumental; Actor in these sentences can be general, and thus it can be omitted on the surface level.

(70) a. *Kniha*$_{Pat}$ *byla napsána slavným autorem*$_{Act}$.
Book$_{Nom}$ was written famous author$_{Ins}$.
'The book was written by a famous author.'

b. *Bazén byl vypuštěn.*
Swimming pool was emptied.

There is another possible surface form of Actor: the prepositional phrase *od* (from) + Genitive, but this form cannot be used with all verbs—here, the semantics of the verb and its participants plays a role:

(71) a. *Pepík je bit od otce.*
Pepík is beaten from father.

b. **Kniha byla napsána od slavného autora.*
Book was written from famous author.

The conditions in which this construction can be used will be examined in the future work. Here, we assume that Actor can only change to Instrumental.

Before we start describing the algorithm, we have to make one more important remark: when we speak about a change of the structure we always work with an *instance* of a verb frame. The verb frame is an abstract set of all possible realizations, and we can only make a diathesis of a chosen member of this set.

The algorithm for deriving the frame of the periphrastic passive is described here:

- The verb form changes to periphrastic passive.
- If there is a nominal object in Accusative in the frame, it becomes subject (in Nominative). The subject marker changes so that it pointed to the new subject.
- If the object in Accusative is a clause or the infinitive, it becomes the subject, with a special sort of agreement (3rd person, singular, neuter).
- If there is no object in Accusative the passive has empty subject, with the same sort of agreement as the infinitive or clause subject. The subject marker is deleted.
- If our frame instance contains only the subject on the surface, this type of passivization is prohibited.
- The original subject becomes a generalizable member which is realized by Instrumental.
- All other members of the frame stay in their positions.

There are some exceptions to the above rules. The first group of exceptional verbs are ditransitive verbs (verbs with two Accusatives in the frame). We have found only two such verbs in Czech:

- *stát koho co* - to cost sb sth

  This verb does not have the passive.

- *učit koho co$_{acc}$/čemu$_{dat}$* - to teach sb sth

  If we choose the frame with Accusative and Dative, no problems occur. But in the frame with two Accusatives, one of them must be omitted (both can be generalized) before we create the passive construction:

(72) a. *Děti jsou učeny (matematice).*
Children are taught (to mathematics$_{Dat}$).

b. **Děti jsou učeny matematiku.*
Children are taught mathematics$_{Acc}$.

c. *Matematika je učena.*
Mathematics is taught.

d. **Matematika je učena děti.*
Mathematics is taught children$_{Acc}$.

The periphrastic passive is marked by `%` in the lexicon, and the entries of the verb *učit* will look as follows:

(73) a. `učit~2` `R--s[i1]1(hPc1)2(hTc3)3(hPc4)%$`

b. `učit~2` `R--s[i1]1[hPc1]2[hTc4]3(hPc4)@`

c. `učit~2` `P--s[i2]1(hPc7)2[hTc1]@`

d. `učit~2` `PSEs[i2]1(hG)2[hTc1]@`

Another exceptional group of verbs are reflexives tantum which can have passive forms. The member of the frame which undergoes the change into subject is not a member in Accusative but in Genitive:

(74) a. *Soudce se tázal svědka, zda něco viděl.*
Judge$_{Nom}$ SE asked witness$_{Gen}$ if he saw anything.

b. *Svědek byl (soudcem) tázán, zda něco viděl.*
Witness$_{Nom}$ was (judge$_{Ins}$) asked if he saw anything.

This group of verbs is not very numerous. It contains verbs *tázat se* (and its prefixed variants), *obávat se*, and perhaps some more. It is a question whether we should introduce new rules for this type of passive or rather store these passive frames as exceptions:

(75) a. `tázat` `RSEs[i1]1[hPc1]2[sF|sR|hPTZc4{na}]3(hPTc2)@`

b. **tázat** `P--s[i3]1(hPc7)2(sF|sR|hPTZc4{na})3[hPTc1]@`

The periphrastic passive is felt as rather formal, bookish or obsolete in modern Czech, especially the passive with expressed Actor. Unlike its English counterpart, Czech passive is rarely used for changing the topic-focus articulation—for this purpose the change of the word order is employed. The passive construction is mainly used, if the speaker wants to avoid saying who/what Actor is, or if Actor is general. In both these cases, however, the reflexive passive is used more often.

4.3.3. Reflexive passive

In this construction, the verb changes its form to reflexive passive form, the participant in Accusative (if present) becomes the subject, and Actor becomes general.

(76) a. *Bábovka se peče.*
Cake SE bakes.
'The cake is being baked.'

b. *Do města se jde tudy.*
To town SE goes this way.
'This is the way to the town.'

The example in (76a) is the real reflexive passive, derived from a transitive verb, while the sentence in (76b) is an impersonal active construction, derived from an intransitive verb. We mark both these constructions as reflexive passive as the algorithms for deriving them are very similar.

The reflexive passive is sometimes indistinguishable from the intrinsic or true reflexive. The sentence

(77) *Děti se učí dobře.*
Children SE teach well.
'Children are easy to teach.' or 'The children learn well.'

has two readings, as the verb *učit* 'to teach' in reflexive passive has the same form as the reflexive verb *učit se* 'to learn'. This ambiguity is inherent in the language and we will not try to solve this problem in the lexicon.

The algorithm for deriving the reflexive passive frame is nearly identical with the algorithm for the periphrastic passive:

- The verb changes its form to a reflexive passive form.
- If there is a nominal object in Accusative in the frame, it becomes subject (in Nominative). The subject marker is changed so that it points to the new subject.

- If the object in Accusative is a clause or the infinitive, it becomes the subject, with a special sort of agreement (3rd person, singular, neuter).
- If there is no object in Accusative the passive has an empty subject, with the same sort of agreement as the infinitive or clause subject. The subject marker is deleted.
- The original subject is generalized (and thus omitted on the morphemic level).
- All other members of the frame stay in their positions.

The rules for handling the ditransitive verbs *stát* 'to cost' and *učit* 'to teach' are the same as at the periphrastic passive: *stát* cannot be passivized and with the verb *učit*, the frame to be passivized can contain only one Accusative (see 73).

(78) a. *Děti se učí (matematice).*
Children$_{Nom}$ SE teach (to mathematics$_{Dat}$).

b. **Děti se učí matematiku.*
Children SE teach mathematics$_{Acc}$.

c. *Matematika / Matematice se učí od první třídy.*
Mathematics$_{Nom/Dat}$ SE teaches from first grade.

d. **Matematika / Matematice se učí děti.*
Mathematics$_{Nom/Dat}$ SE teaches children$_{Acc}$.

The reflexive passive of *učit*, however, is homonymous with the reflexive verb *učit se* 'to learn', and thus it is difficult for a Czech speaker to understand the examples in (78a) and (78b) in the passive meaning. As an active sentence with the verb *učit se*, (78b) is correct.

Reflexive passive is marked by `$` in the lexicon and an example of a lexical entry was given in (73a).

For the proper treatment of the verb *učit* we also have to add an irregular frame for the reflexive passive:

(79) `učit~2          PSEs[i2]1(hG)2[hTc1]@`

The reflexive passive is used especially in cases when Actor is general and the periphrastic passive cannot be used:

(80) a. *Tady se hodně čte.*
Here SE much reads.
'Here, people read a lot.'

b. **Tady je hodně čteno.*
Here is much read$_{Prtcpl}$.

c. *Matematice se učí od první třídy.*
Mathematics$_{Dat}$ SE teaches from first grade.

d. ?*Matematice je učeno od první třídy.*
Mathematics$_{Dat}$ is taught from first grade.

4.3.4. Mediopassive

This construction is very similar to the previous one—some linguistic books actually do not distinguish between them. In mediopassive, Actor is present (though it can be general) and an adverb like *dobře* (well), *špatně* (badly), *snadno* (easily), etc. (i.e. free modification of Manner), must be present in the construction. This type of passive was described by M. Dokulil (1941) as a special case of description of the way something is done. P. Karlík (1995) considers this construction a special case of the subject diathesis of the type agent–patient where the agentive role is put to the background and the agent is getting a role of experiencer.

Examples:

(81) a. *Matematika se mi učí snadno.*
Mathematics$_{Nom}$ SE me$_{Dat}$ learns easily.
'It's easy for me to learn/teach mathematics.'

b. *Z této látky se šije dobře.*
From this fabric SE sews well.
'It's easy (for anyone) to make clothes from this fabric.'

This construction can also be ambiguous—either with a reflexive passive or with an intrinsic passive. The Dative member is then understood as Benefactor:

(82) a. *Děti se mi učí dobře.*
Children SE to me teach well.
'It's easy for me to teach children.' or 'My children learn well.'

b. *Teď už se mi píše potvrzení dobře.*
Now already SE me$_{Dat}$ writes receipt$_{Nom}$ well.
'Now, the receipt is finally being written correctly for me.' or
'Now, it's already easy for me to write the receipt.'

The mediopassive can also be derived from an intransitive verb:

(83) *S kopce dolů se (mi) jde dobře.*
From hill down SE (me$_{Dat}$) goes well.
'It's easy (for me) to walk down-hill.'

The algorithm for deriving the mediopassive frame is nearly identical with the algorithm for the periphrastic passive:

- The verb form is changed in a reflexive passive form.
- If there is a nominal object in Accusative in the frame, it becomes subject (in Nominative). The subject marker is changed so that it points to the new subject.
- If the object in Accusative is a clause or the infinitive, it becomes the subject, with a special sort of agreement (3rd person, singular, neuter).
- If there is no object in Accusative the passive has an empty subject, with the same sort of agreement as the infinitive or clause subject. The subject marker is deleted.
- The original Actor (subject) changes its surface realization to Dative.
- All other members of the frame stay in their positions.

We do not introduce a separate mark for the possibility of deriving mediopassive as we believe that there is a general rule: any frame of an imperfective verb which can be transformed to reflexive passive can also be transformed to mediopassive. The information on reflexive passives is containe in our lexicon, and the information on aspect is contained in the morphological lexicon. If it turned out that the above rule does not hold we can introduce a new mark.

There is, however, a verb that needs special treatment: the verb *chtít* can have a reflexive form *chtít se* where Actor has the form of Dative. We will call this construction mediopassive, but it requires a separate entry in the lexicon. As this verb requires an infinitive in its frame we will show the encoding of the frame in Section 4.4.2.

4.3.5. Constructions with *mít* and *dostat*

In this type of construction, a Dative member of the frame (typically Addressee) becomes the subject of a construction with the support verb *mít* or *dostat* and the main verb occurs in the predicate as a passive participle in Accusative. If the main verb has an Accusative object (typically Patient), the participle agrees with it in gender and number. If the Accusative object is missing, the participle has the form of singular neuter. Actor (the original subject) becomes an optional member of the frame in the form of *od* + Genitive:

(84) a. *Obec přidělila žadatelům byty.*
Municipality$_{Nom}$ granted applicants$_{Dat}$ flats$_{Acc}$.

b. *Žadatelé mají/dostali (od obce) přiděleny byty.*
Applicants$_{Nom}$ have/got (from municipality) granted$_{Prtcpl\,Acc}$ flats$_{Acc}$.
'Applicants were granted flats (by municipality).'

c. *Otec$_{Act}$ vynadá Pepíkovi$_{Pat}$.*
Father$_{Nom}$ will scold Pepík$_{Dat}$.

d. *Pepík$_{Pat}$ dostane vynadáno (od otce$_{Act}$).*
Pepík$_{Nom}$ will get scolded (from father$_{Gen}$).
'Pepík will be scolded (by the father).'

e. *Vnučka babičce uvařila.*
Granddaughter$_{Nom}$ grannie$_{Dat}$ cooked.
'Granddaughter has cooked for grannie,'

f. *Babička má uvařeno.*
Grannie$_{Nom}$ has cooked$_{Prtcpl\,Neut\,Sg}$.
'(The meal) has been cooked for grannie.'

Some verbs allow both of the two support verbs, while others allow only one of them (*mít/dostat přiděleno*, *dostat/*mít vynadáno*, **dostat/mít uvařeno*). This is why we introduced two marks—one for each of the support verbs. The frames

Instead of the (short) passive participle we can use the long form of adjective (long passive participle), especially in the spoken language. In such a case, however, the sentence can become ambiguous:

(85) a. *Žadatelé dostali (od obce) přiděleny byty.*
Applicants got (from municipality) granted$_{Prtcpl\,Acc}$ flats$_{Acc}$.
'Applicants were granted flats (by municipality).'

b. *Žadatelé dostali (od obce) přidělené byty.*
Applicants got (from municipality) granted$_{Adj\,Acc}$ flats$_{Acc}$.
'Applicants were granted flats (by municipality).' or
'Applicants got the granted flats (from municipality).

The algorithm for deriving the verb frame of this construction follows:

- An object in Dative (Addressee, Patient, or Beneficiary) becomes subject (in Nominative). The subject marker is changed accordingly.
- Actor becomes an optional member of the frame of the form *od* + Genitive.
- All other members of the frame stay in their positions.

Frames of the verbs which allow this diathesis are in the following example (the diathesis with *mít* is marked by # and the diathesis with *dostat* is marked by *):

(86) a. `přidělit` `R--s[i1]1(hPc1)2[hPTc4]3[hPc3]%$#*`

b. `vynadat` `R--s[i1]1(hPc1)2[hPc3]%$*`

c. `uvařit` `R--s[i1]1(hPc1)2[hPTc4]3<hPc3>%$#~`

4.3.6. Resultative construction with *mít*

There is one more construction with the support verb *mít*. This is not really a diathesis, as Actor remains as subject and the change on the surface only affects the verb form. It is rather a sort of resultative tense, which corresponds to English perfective constructions. K. Hausenblas (1963) ranks this construction to verb tense, while J. Panevová (1971) considers it a sort of aspect. We decided to include this construction among other diatheses because we have no other means how to create these constructions.

(87) a. *Upeču bábovku.*
Bake$_{1SgFut}$ cake$_{FemAcc}$.

b. *Bábovku už mám upečenu/upečenou.*
Cake$_{FemAccSg}$ already have$_{1Sg}$ baked$_{PrtcplFemAccSg\,/\,AdjFemAccSg}$.

c. *Už mám upečeno.*
Already have$_{1Sg}$ baked$_{PrtcplNeutAccSg}$.

In this derivation, the frame remains the same as in the base form. The only operation in forming this construction is changing the predicate.

All the above constructions can only be derived from perfective verbs, as they express a result.

This diathesis is marked by ~, and an example of a verb frame allowing this diathesis is in (86c).

4.4. Verbs with the infinitive in their frames

For this group of verbs, we have to describe not only the frame of the verb, but also the interaction between the higher verb and the lower verb (the infinitive)—which members of the frames they share, what kinds of derived frames are allowed for both the infinitive and the governor, and other constraints that hold for both the verbs.

These verbs are usually divided into two subclasses: **raising** and **equi** (or **control**) verbs. In both cases the subject (or rarely an object) of the infinitive is the subject or an object of the higher verb, but there is a difference between the two deep structures.

Raising verb: The subject of the infinitive becomes (is raised as) the subject or an object of the governor, but it does not belong among its arguments.

Equi verb: Certain participant of the governor is coreferential with a participant of the dependant. On the surface level, such a participant is present only once, but in the deep structure, there are two slots (one in every verb's frame) which are corefential.

Many authors were concerned with these sorts of verbs; this topic is worked up well for English (see Chomsky, 1986; Dalrymple et al., 1995; Pollard and Sag, 1994), for Czech, we will proceed from Panevová (1996) but our conclusions will be different in some cases.

First we will show the difference between the two types of verbs in examples of tree structures. We will use one raising verb (*zdát se*—seem) and one equi verb (*snažit se*—try) for explanation.[15]

(88) a.

- *zdát_se*
  - $GNRL_{Act}$
  - $být_{Pat}$
    - $Jirka_{Act}$
    - $veselý_{Pat}$

Jirka se zdá být veselý.
Jirka SE seems be$_{Inf}$merry.

b.

- *snažit_se*
  - $Jirka_{Act,i}$
  - $být_{Pat}$
    - $COR_{Act,i}$
    - $veselý_{Pat}$

Jirka se snaží být veselý.
Jirka SE tries be$_{Inf}$merry.

For English, certain tests were developed which should show whether a verb is raising or equi. We will examine them and check whether these or similar criteria can be used for Czech as well.

Here are conditions for subject-raising verbs. Conditions and examples are taken from Borsley (1999).

- Raising verbs can have a dummy subject (i.e. an expletive pronoun or *there*), while equi verbs cannot:

  (89) a. It seemed to be easy to please Rhodes.
  b. There seemed to be a flaw in the argument.
  c. *It tried to be easy to please Rhodes.
  d. *There tried to be a flaw in the argument.

[15] J. Panevová (1996) claims that raising verbs are marginal in Czech. She compares Czech with Russian, and though she shows a different behaviour of the verb *zdát se* (seem) in Czech, she more or less adopts Comrie's conclusion (Comrie, 1991) that it is not justified to speak about *Sb-to-Sb raising* in Czech. We will come to different conclusions, however.

This test is hardly applicable for Czech, as Czech only rarely uses dummy subjects. However, the dummy subjects in the above sentences stand for the infinitives, so we can try corresponding Czech sentences with infinitives in the subject positions:

(90) a. *Potěšit Karla se zdá být snadné.*
‘To please Karel seems to be easy.’

b. *Potěšit Karla musí být snadné.*
‘To please Karel must be easy.’

c. *Potěšit Karla začne být snadné.*
‘To please Karel starts to be easy.’

d. **Potěšit Karla zkouší být snadné.*
‘To please Karel tries to be easy.’

e. **Potěšit Karla chce být snadné.*
‘To please Karel wants to be easy.’

In the above examples, the raising verb *zdát se* has two actants (general Actor in Dative and Patient), and the verb *začít* has only one actant (Actor); the subject of the upper verb is raised from the frame of the construction *být snadné* (be easy). Let us show it on graphs of sentences (90a) and (90c):

(91) a.

zdát_se
$GNRL_{Act}$ *být_snadné*$_{Pat}$
potěšit$_{Act}$
$GNRL_{Act}$ *Karla*$_{Addr}$

b.

začít
být_snadné$_{Act}$
potěšit$_{Act}$
$GNRL_{Act}$ *Karla*$_{Addr}$

Modal verbs like *muset* (must) are treated as grammatemes in FGD and therefore they cannot have any actants. A trivial corollary of this fact is that the subject of the modal verb *must* be actant of the infinitive.

(92) a.

být_snadné$_{debitive}$
potěšit$_{Act}$
$GNRL_{Act}$ *Karel*$_{Addr}$

Potěšit Karla musí být snadné.
Please$_{Inf}$Karel must be$_{Inf}$easy.

b. *být*$_{debitive}$ — *Jirka*$_{Act}$ *veselý*   *Jirka musí být veselý.*
Jirka must be$_{Inf}$merry.

- Raising verbs can have a clausal subject, while equi verbs cannot:

(93) a. That he is clever seems to be obvious.
b. *That he is clever tries to be obvious.

Let us check whether this test works for Czech:

(94) a. *Že je chytrý, se zdá být zřejmé.*
'That he is clever seems to be obvious.'
b. *Že je chytrý, musí být zřejmé.*
'That he is clever must be obvious.'
c. *Že je chytrý, začne být zřejmé.*
'That he is clever starts to be obvious.'
d. **Že je chytrý, zkouší být zřejmé.*
'That he is clever tries to be obvious.'
e. **Že je chytrý, chce být zřejmé.*
'That he is clever wants to be obvious.'

- Idiomatic expressions keep their idiomatic meaning in the raising sentences, but not in the control sentences:

(95) a. The cat seems to be out of the bag. (idiomatic meaning)
b. The cat tries to be out of the bag. (non-idiomatic meaning)

(96) a. *Už je ruka v rukávě.*
Already is arm in sleeve.
'Everything has been arranged.'
b. *Ruka se zdá být v rukávě.* (idiomatic meaning)
c. *Ruka musí být v rukávě.* (idiomatic meaning)
d. ?*Ruka začne být v rukávě.* (sounds odd)
e. ?*Ruka zkouší být v rukávě.* (sounds odd)
f. *Ruka chce být v rukávě.* (sounds odd)

This test does not seem to work for Czech, but there are some more.

The following conditions and examples are taken from Pollard and Sag (1994).

- Object-raising verbs can have a dummy subject (like the subject-raising verbs):

(97) a. Kim believed there to be some misunderstanding about these issues.

b. *Kim persuades there to be some misunderstanding about these issues.

Such constructions are impossible in Czech, and thus we cannot make a similar test.

- Controllers of equi constructions have assigned semantic roles:

(98) a. The doctor tried to examine Sandy.

b. Sandy tried to be examined by the doctor.

The doctor is Actor of the verb try in (98a) ("tryer" in HPSG terminology), while in (98b), Sandy is Actor.

In Czech, we can find similar examples:[16]

(99) a. *Doktor se pokusil vyšetřit babičku.*
Doctor SE tried examine$_{Inf}$ grannie.

b. *Babička se pokusila nechat se vyšetřit od doktora.*
Grannie SE tried let$_{Inf}$ SE examine$_{Inf}$ from doctor.

- Another test examines the meaning of the active and passive infinitive. In raising sentences, the sentences with active and passive infinitives are identical with respect to their contents:

(100) a. Kim believed the doctor to have examined Sandy. (raising verb)

b. Kim believed Sandy to have been examined by the doctor. (raising verb)

c. Kim persuaded the doctor to examine Sandy. (equi verb)

d. Kim persuaded Sandy to be examined by the doctor. (equi verb)

In Czech, this test can be applied, too:

(101) a. *Karel viděl doktora vyšetřit babičku.*
Karel saw doctor examine$_{Inf}$ grannie.

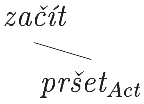
b. *Karel viděl babičku nechat se vyšetřit od doktora.*
Karel saw grannie let$_{Inf}$ SE examine$_{Inf}$ from doctor.

c. *Karel nařídil doktorovi vyšetřit babičku.*
Karel ordered doctor examine$_{Inf}$ grannie.

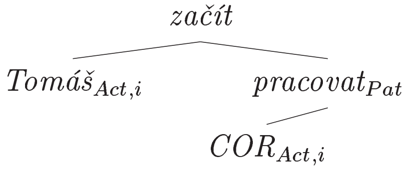
d. *Karel nařídil babičce nechat se vyšetřit od doktora.*
Karel ordered grannie let$_{Inf}$ SE examine$_{Inf}$ from doctor.

[16]It is impossible to use periphrastic passive in these constructions, the only possibility how to paraphrase the English sentences is to use a support verb *nechat* (let), and we can understand the construction as a sort of passive.

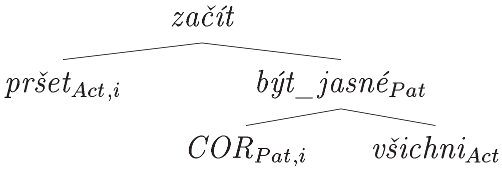
Another test, which can be applied, checks the number of actants of the upper verb, and their surface realization. This number should not depend on the lexical setting of the infinitive. And also the surface realization of a certain actant should not depend on the lexical setting of another actant. If we considered for example that the verb *začít* (start) is an equi verb whose subject is coreferential with the subject of the embedded infinitive we would need several frames:

(102) a.


Začalo pršet.
Started rain$_{Inf}$.

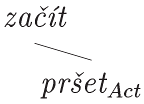
b.


Tomáš začal pracovat.
Tomáš started work$_{Inf}$.

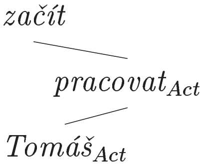
c.


Že prší, začalo být jasné všem.
That rains started be$_{Inf}$clear all$_{Dat}$.

The verb from (102a) would have a frame with Actor realized as an infinitive. The verb from (102b) would have a frame with Actor realized by a noun in Nominative and Patient realized by an Infinitive. The verb from (102c) would have also Actor and Patient in its frame, but Actor would be ralized by a clause attached by *že*. We can see that we could continue and find even more different frames for the equi verb *začít*. On the other hand, if we suppose that the verb *začít* is a raising verb we get rid of the problem with many frames. The frame only contains Actor (the infinitive) and the subject is raised from Actor's frame. It can be even empty if the infinitive has no subject.

(103) a.


Začalo pršet.
Started rain$_{Inf}$.

b.


Tomáš začal pracovat.
Tomáš started work$_{Inf}$.

c. *začít*

být_jasné$_{Act}$

debouncing$_{Pat}$ *všichni*$_{Act}$

Že prší, začalo být jasné všem.
That rains started be$_{Inf}$clear all$_{Dat}$.

We used a similar consideration for the so-called Slavic Accusative in sentences with verbs of perception. We believe that sentences in (104) have identical content:

(104) a. *Petr viděl doktora vyšetřit babičku.*
Petr saw doctor$_{Acc}$ examine$_{Inf}$ grannie$_{Acc}$.

b. *Petr viděl doktora, jak vyšetřuje babičku.*
Petr saw doctor$_{Acc}$ how examines grannie$_{Acc}$.

c. *Petr viděl, jak doktor vyšetřuje babičku.*
Petr saw how doctor$_{Nom}$ examines grannie$_{Acc}$.

The verb *vidět* has only two actants, in our model, and the above sentences could be expressed by the structure in (105):

(105) *vidět*

Petr$_{Acc}$ *vyšetřit*$_{Pat}$

doktor$_{Act}$ *babička*$_{Pat}$

Now, we have tools for judging equi and raising verbs and we can start describing single lexical entries.

4.4.1. Raising verbs

First, we will deal with **subject raising verbs**. This group of verbs contains mainly the modal and aspectual verbs. As it was said above, modal verbs are considered grammatemes in FGD and thus they cannot have own argument structure. On the surface level, however, they impose certain constraints on the infinitives. These constraints must be encoded in the lexicon and that is why we introduce lexical entries for these verbs.[17]

In examples in (106) we show various constructions of raising verbs; the members of the infinitival clauses are enclosed in brackets and the trace of the raised element is marked by an underscore.

[17] There exist also other auxiliaries which need entries in the lexicon, e.g. prepositions, which impose a certain morphological case on the following nominal group.

(106) a. *Petr$_{Act,i}$ smí [_$_i$ odejít].*
Petr$_{Nom}$ may leave$_{Inf}$.

b. *Začalo [pršet].*
Started rain$_{Inf}$.
'It started raining.'

c. *Petr$_{Pat,i}$ musí [_$_i$ být pochválen].*
Petr$_{Nom}$ must be$_{Inf}$ praised$_{Prtcpl}$.

d. *Musí [_$_i$ se zabít] dvě mouchy$_{Pat,i}$ [jednou ranou].*
Must SE kill$_{Inf}$ two flies$_{Nom}$ one hit$_{Ins}$.
'Two flies must be killed by one hit.'

e. *Bábovka$_{Pat,i}$ [se] začala [_$_i$ péci].*
Cake$_{Nom}$ SE started bake$_{Inf}$.
'The cake started to be baked.'

f. *Únosce$_{Addr,i}$ musí [_$_i$ dostat slíbeno výkupné$_{Pat}$].*
Kidnapper$_{Nom}$ must get$_{Inf}$ promised$_{Prtcpl}$ ransom$_{Acc}$.
'The kidnapper must be promised the ransom.'

g. *Kuchařka$_{Act,i}$ [už] musí [_$_i$ mít uvařeno].*
Cook$_{Nom}$ already must have$_{Inf}$ cooked$_{PrtcplNeutSg}$.
'The cook must have already cooked (everything).'

h. *[Tady se ti$_{Act}$] musí [sedět nepohodlně].*
Here SE you$_{Dat}$ must sit$_{Inf}$ uncomfortably.
'This must be an uncomfortable seat for you.'

We can see in the above examples that the infinitive can occur in various diatheses. The infinitive can occur in both periphrastic and reflexive passive and in the construction with the verb *dostat*; the mediopassive and the active construction with the verb *mít* are only possible with the verb *muset* (must) in the meaning of high probability. It seems that the governor can only occur in active voice, but we will come back to this issue later.

As modal verbs have no representation on the tectogrammatical level we have to find a notation of these lexical entries that respects this theoretical constraint and gives all necessary information. In (107) we can see several examples of both modal and non-modal verbs.

(107) a. `muset~1` `R--s[a0]0[sId%$#mD]@`

b. `muset~2` `R--s[a0]0[sId%$*~]@`

c. `začít` `R--s[a1]1[sId%$]@`

d. `zdát` `RSEs[i2]1(hPc3)2[hTc1|sD]@`

e. `zdát` `RSEs[a2]1(hPc3)2[sIl{být}|hQc1]@`

The frame of the modal verb *muset* (107a) contains only one "argument" (0[sId%$ mD]) whose functor is marked by 0 (zero). This notation was adopted for sentence complementations which do not belong to frame of a given verb. Attributes enclosed in brackets represent constraints imposed on the surface forms. In (107a) these attributes have the following values: infinitive (sI) which can occur in periphrastic and reflexive passive (d%$),[18] and the modality feature *debitive* (mD). The subject of the construction is raised from the infinitival clause (s[a0]). The verb *muset* can occur only in active voice (@).

The frame of the verb *muset* in the meaning of high probability (107b) is very similar the frame of the modal verb. It differs in constraints imposed of diatheses of the embedded infinitive (%$*~) and in a missing modality marker.

The aspectual (phase) verb *začít* is a verb with one actant (Actor: 1[sId%$]) which is realized by an infinitive. The infinitive can occur also in periphrastic or reflexive passive, and the verb *začít* can only occur in active voice. The subject of the verb *začít* is raised from the infinitival clause.

The verb *zdát se* has been already discussed. In (107d) we can see the frame of the verb with

Object raising verbs are such verbs that have an infinitive in the frame and the subject of this infinitive becomes an object of the higher verb. This group contains the verbs of perception:

(108) a. *Vidím* ho_i $\_\_i$ *přicházet.*
I see him to be coming.
'I see him coming.'

b. ?*Vidím* ho_i $\_\_i$ *být tázána.*
?I see him to be asked.
'I see him being asked.'

c. ?*Cítím* $bábovku_i$ $\_\_i$ *péct se.*
?I smell cake to bake SE.
'I can smell that a cake is being baked.'

The passive constructions are questionable with this group of verbs; a further research on a text corpus will be necessary. In the current version of the lexicon the possibility of creating the passive voice is suppressed. The frame is encoded this way:

[18] These constraints represent additional constraints to those imposed by the lexical entry of a given infinitive.

(109) a. `vidět` `R--s[i1]1(hPc1)2(hPTSc4|sD|sZ)&`

b. `vidět` `R--s[i1]1[hPc1]2[sId$|sZd&]0[hPTSc4a2]@`

For marking the source of the raised subject we use the attribute *a*. Its value points to a functor from which the subject was raised.

The last issue that will be discussed in this section is the possible reflexivization of modal and aspectual verbs. As we have said above, raising verbs do not seem to allow passivization, but let us consider the following conversation:

(110) a. *Honza/*Honzu se musí požádat o povolení.*
Honza$_{Nom/*Acc}$ SE must ask$_{Inf}$ for permission.
'Honza must be asked for permission.'

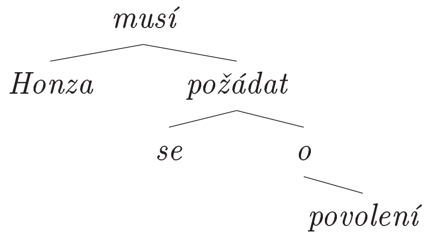
b. *Co že se musí udělat?*
What$_{Nom}$ that SE must do$_{Inf}$?
'What did you say that must be done?'

c. *Požádat Honzu o povolení.*
Ask$_{Inf}$ Honza$_{Acc}$ for permission.

d. *Žádat*$_{Inf}$ *Honzu o povolení se mi nechce.*
Ask$_{Inf}$ Honza$_{Acc}$ for permission SE me$_{Dat}$ wants$_{Neg}$.
'I don't want to ask Honza for permission.'

e. *Požádat Honzu/*Honza o povolení se musí!*
Ask$_{Inf}$ Honza$_{Acc/*Nom}$ for permission SE must!
'Honza MUST be asked for permission!'

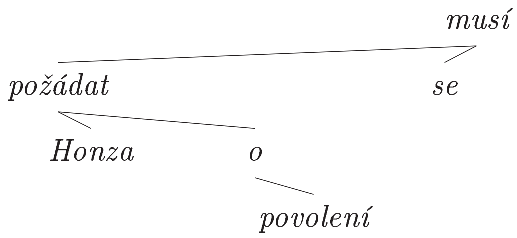
In the sentence (110a) the embedded infinitive is in reflexive passive and its subject (Addressee in the deep structure) is raised as the subject of the modal verb muset. In (110e) the infinitive is in active voice, with Addressee in Accusative. The whole infinitive clause is the subject of reflexive passive of the verb *muset*. We will show the difference on structures on the analytical level:[19]

(111) a.


Honza se musí požádat o povolení.
Honza$_{Nom}$ SE must ask$_{Inf}$ for permission.

[19]The analytical level is an auxiliary level where analytical morphological forms are resolved. This level is used for example in Prague Dependency Treebank (PDT) as one of the levels of description of the texts (see Hajič et al., 1999).

b.



Požádat Honzu o povolení se musí.
Ask_{Inf} Honza_{Acc} for permission SE must.

The (exceptional) frame for the reflexive passive looks as follows:

(112) `muset~1      PSEs[i0]0[sId@mD]@`

4.4.2. Equi verbs

This type of verbs in Czech was described by K. Svoboda (1962) and J. Panevová (1996). Svoboda does not use the term *equi* or *control*, but he distinguishes between "subject infinitives" verb and "object infinitives". He does not distinguishes raising and equi verbs, as he only considers the surface structure and grammatical functions as subject, objects, etc.

Panevová describes carefully equi verbs from the point of view of FGD. She distinguishes four types of equi verbs:

(113) a. Subject-control (Act-Sb):
$\textit{Jan}_{Act,i}$ se bojí [$\_\_{}_{Act,i}$ zůstat doma sám].
Jan_{Nom} fears stay$_{Inf}$ at home alone.

b. Object-control (Addr-Sb):
$\textit{Oni}_{Act}$ $\textit{mu}_{Addr,i}$ poručili [$\_\_{}_{Act,i}$ přijít].
They_{Nom} him_{Dat} ordered come_{Inf}.

c. Ambiguous class (Act-Sb) or (Addr-Sb):
$\textit{Rodiče}_{Act,i}$ $\textit{Petrovi}_{Addr,j}$ slíbili [$\_\_{}_{Act,j}$ svézt se na poníkovi].
Parents to Petr promised to ride on pony.
$\textit{Rodiče}_{Act,i}$ $\textit{Petrovi}_{Addr,j}$ slíbili [$\_\_{}_{Act,i}$ přestat kouřit].
Parents to Petr promised to stop to smoke.

d. Object-control (Pat-Sb) (the infinitive has the function of Intent):
Sedlák vyhnal $\textit{krávy}_{Pat,i}$ [$\_\_{}_{Act,i}$ pást se].
Farmer drove cows_{Acc} graze_{Inf}.

We will add two more types, which are quite rare but interesting. The embedded infinitive should be understood as a sort of passive, though it is in active voice:

(114) a. (Act-Addr) control:
Anežka$_{Act,i}$ *chce* [$_$$_{Act}$ *podat* *knihu* $\_$$_{Addr,i}$].
Anežka$_{Nom}$ wants pass$_{Inf}$ book$_{Acc}$.
'Anežka wants someone to pass her the book.'
Anežka$_{Act,i}$ *chce* [$_$$_{Act}$ *přečíst* *pohádku* $\_$$_{Addr,i}$].
Anežka wants read$_{Inf}$ tale$_{Acc}$.
'Anežka wants someone to read her a tale.'
Anežka$_{Act,i}$ *chce* [$_$$_{Act}$ *poučit* *o hudbě* $\_$$_{Addr,i}$].
Anežka wants to instruct in music.
'Anežka wants someone to instruct her in music.'

b. (Act-Pat) control:
Plot$_{Act,i}$ *chce* [$_$$_{Act}$ *natřít* $\_$$_{Pat,i}$].
Fence wants paint$_{Inf}$.
'The fence needs painting.'
Pepík$_{Act,i}$ *potřebuje* [$_$$_{Act}$ *nařezat* $\_$$_{Pat,i}$].
Pepík needs spank$_{Inf}$.
'Pepík needs spanking.'

For proper description of all the above constructions in the lexicon we also have to examine the possible diatheses of both the governer and the controlled infinitive. Let us start with (Act-Sb) control:

(115) a. *Petr*$_{Act,i}$ *chce* $_$$_{Pat,i}$ *být* *pochválen.*
Petr wants to be praised.

b. *Anežka*$_{Act,i}$ *chce* $_$$_{Addr,i}$ *být* *poučena* *o hudbě.*
Anežka wants to be instructed in music.

c. *Bábovka*$_{Act,i}$ *se* *nechce* $_$$_{Pat,i}$ *péct.*
Cake SE does not want to bake.
'The cake refuses to get baked.'

d. *Pepík*$_{Act,i}$ *nechce* $_$$_{Pat,i}$ *dostat nařezáno.*
Pepík does not want to get spanked.
'Pepík does not want to be spanked.'

e. *Petr*$_{Act,i}$ *chce* $_$$_{Pat,i}$ *dostat/*mít* *slíbenu* *hračku.*
Petr wants to get/*have promised toy.
'Petr wants to be promised a toy.'

f. *Matka*$_{Act,i}$ *už chce* $\_{}_{Act,i}$ *mít uvařeno.*
Mother already wants to have cooked.
'Mother wants to have all cooking done already.'

We can see that the infinitive can be in passive, as well as in a construction with *mít* or *dostat*. The passivization of the governor, on the other hand, does not seem to be possible. The reason may be that the subject of the embedded infinitive is controlled by Actor which would become general in a passive construction. An exception is a mediopassive of the verb *chtít*.

(116) *Nechce se mi*$_{Act,i}$ $\_{}_{Act,i}$ *spát.*
Wants$_{Neg3SgNeut}$ SE me$_{Dat}$ sleep$_{Inf}$.
'I don't want to sleep'.
Nechce se mi$_{Act,i}$ $\_{}_{Pat,i}$ *být bit.*
Wants$_{Neg3SgNeut}$ SE me$_{Dat}$ be$_{Inf}$ beaten.
'I don't want to be beaten'.
Bábovce$_{Act,i}$ *se nechce* $\_{}_{Pat,i}$ *péct (se).*
Cake$_{Dat}$ SE wants$_{Neg3SgNeut}$ bake$_{Inf}$ (SE).
'The cake refuses to get baked'.

The verb *chtít* even allows reflexive passive with general Actor:

(117) *Když se* $\_{}_{Act,i}$ *nechce* $\_{}_{Act,i}$ *pracovat, tak se nemusí* $\_{}_{Act,i}$ *jíst.*
When SE wants$_{Neg3SgNeut}$ work$_{Inf}$ then SE needs$_{Neg}$ eat$_{Inf}$.
'If one doesn't want to work then he doesn't need to eat.'

Frames of two equi verbs, *bát se* (fear) and *chtít* (want) follow:

(118) a. `bát` `RSEs[i1]1[hPc1]2(hPTRc2|hPTRc4r{o}|sD|sU|sIq1d%)@`

b. `chtít~1` `R--s[i1]1[hPc1]2[hTc4|sIq1d%$#~]@`

c. `chtít~2` `PSEs[i2]1(hPTc3)2[hZc4|sIq1d%$]@`

Next, we will examine the the possibility of passivization of verbs with (Pat-Sb) control.

(119) a. *Velitelé*$_{Act,i}$ *vojákům*$_{Addr,j}$ *zakázali* $\_{}_{j}$ *chodit na pivo.*
Commanders soldiers$_{Dat}$ prohibited go$_{Inf}$ for beer.

b. *Vojákům*$_{Addr,j}$ *bylo (veliteli*$_{Act,i}$*) zakázáno* $\_{}_{j}$ *chodit na pivo.*
Soldiers$_{Dat}$ was (commanders$_{Ins}$) prohibited go$_{Inf}$ for beer.

c. *Vojákům*$_{Addr,j}$ *se zakázalo* $\_\_{j}$ *chodit na pivo.*
Soldiers$_{Dat}$ SE prohibited go$_{Inf}$ for beer.

d. *Vojáci*$_{Addr,j}$ *mají/*dostali (od velitelů*$_{Act,i}$*) zakázáno* $\_\_{j}$ *chodit na pivo.*
Soldiers$_{Nom}$ have/*got (from commanders) prohibited go$_{Inf}$ for beer.

e. *Šéf*$_{Act,i}$ *zabránil podřízenému*$_{Addr,j}$ $\_\_{j}$ *být povýšen.*
Boss prevented employee$_{Dat}$ be$_{Inf}$ promoted.

f. *Podřízenému*$_{Addr,j}$ *bylo (šéfem*$_{Act,i}$*) zabráněno* $\_\_{j}$ *být povýšen.*
Employee$_{Dat}$ was (boss$_{Ins}$) prevented be$_{Inf}$ promoted.

g. *Podřízenému*$_{Addr,j}$ *se zabránilo* $\_\_{j}$ *být povýšen.*
Employee$_{Dat}$ SE prevented be$_{Inf}$ promoted.

Frames for the verbs *poručit* (order), *zakázat* (forbid) and *zabránit* (prevent) follow:

(120) a. `poručit` `R--s[i1]1(hPc1)2[sU|sIq3d@]3(hPc3)%$#`

b. `zakázat` `R--s[i1]1(hPc1)2[sU|sIq3d@]3(hPc3)%$#`

c. `zabránit~1` `R--s[i1]1(hPc1)2[sU|sIq3d%]3[hPc3]%$`

The next category to be examined are the ambiguous verbs like *slíbit* (promise) or *odepřít* (refuse). First, we will examine possible diatheses of the governor.

(121) a. ?*Rodiče*$_{Act,i}$ *Petrovi*$_{Addr,j}$ *slíbili* $\_\_{j}$ *svézt se na poníkovi.*
Parents$_{Nom}$ Petr$_{Dat}$ promised ride$_{Inf}$ on pony.

b. *Petrovi*$_{Addr,j}$ *bylo (rodiči*$_{Act,i}$*) slíbeno* $\_\_{j}$ *svézt se na poníkovi.*
Petr$_{Dat}$ was (parents$_{Ins}$) promised ride$_{Inf}$ on pony.

c. *Petrovi*$_{Addr,j}$ *se slíbilo* $\_\_{j}$ *svézt se na poníkovi.*
Petr$_{Dat}$ SE promised ride$_{Inf}$ on pony.

d. *Petr*$_{Addr,j}$ *má/dostal (od rodičů*$_{Act,i}$*) slíbeno* $\_\_{j}$ *svézt se na poníkovi.*
Petr has/got (from parents) promised ride$_{Inf}$ on pony.

e. *Rodiče*$_{Act,i}$ *Petrovi*$_{Addr,j}$ *slíbili* $\_\_{i}$ *přestat kouřit.*
Parents$_{Nom}$ Petr$_{Dat}$ promised stop$_{Inf}$ to smoke.

f. **Petrovi*$_{Addr,j}$ *bylo (rodiči*$_{Act,i}$*) slíbeno* $\_\_{i}$ *přestat kouřit.*
Petr$_{Dat}$ was (parents$_{Ins}$) promised stop$_{Inf}$ smoke$_{Inf}$.

g. (*)*Petrovi*$_{Addr,j}$ *se slíbilo* $\_\_{i}$ *přestat kouřit.*
Petr$_{Dat}$ SE promised stop$_{Inf}$ to smoke.

h. **Petr*$_{Addr,j}$ *má/dostal (od rodičů*$_{Act,i}$*) slíbeno* $\_{i}$ *přestat kouřit.*
Petr$_{Nom}$ has/got (from parents) promised stop$_{Inf}$ smoke$_{Inf}$.

The construction (121a) is rejected by some speakers, but it can be converted into passive constructions (121b)–(121d), which are admitted by all speakers. The sentence (121e) is perfectly correct, but the passivization of the controller is impossible. Only the sentence in (121g) can be accepted if we suppose Actor of the embedded infinitive to be general.

Let us now try to passivize the infinitive:

(122) a. *Rodiče*$_{Act,i}$ *Petrovi*$_{Addr,j}$ *slíbili* $\_{j}$ *být pochválen.*
Parents$_{Nom}$ Petr$_{Dat}$ promised be$_{Inf}$ praised.

b. *Petrovi*$_{Addr,j}$ *bylo (rodiči*$_{Act,i}$*) slíbeno* $\_{j}$ *být pochválen.*
Petr$_{Dat}$ was (parents$_{Ins}$) promised be$_{Inf}$ praised.

c. *Petrovi*$_{Addr,j}$ *se slíbilo* $\_{j}$ *být pochválen.*
Petr$_{Dat}$ SE promised be$_{Inf}$ praised.

d. *Petr*$_{Addr,j}$ *má/*dostal (od rodičů*$_{Act,i}$*) slíbeno* $\_{j}$ *být pochválen.*
Petr has/*got (from parents) promised be$_{Inf}$ praised.

e. *Rodiče*$_{Act,i}$ *Petrovi*$_{Addr,j}$ *slíbili* $\_{i}$ *být v práci povýšeni.*
Parents$_{Nom}$ Petr$_{Dat}$ promised be$_{Inf}$ at work promoted.

Now, we can encode the frames of the verb *slíbit* (promise):

(123) a. `slíbit~1` `R--s[i1]1(hPc1)2[hZc4|sD|sIq3d%]3[hPc3]%$#*`

b. `slíbit~2` `R--s[i1]1(hPc1)2[hTc4|sD|sIq1d%]3(hPc3)@`

The constructions with (Act-Pat) control and (Act-Addr) control do not allow any sort of diathesis, so their frames will be quite simple:

(124) a. `chtít~3` `R--s[i1]1[hPTc1]2[sIp1d@]@`

b. `chtít~4` `R--s[i1]1[hPTc1]2[sIt1d@]@`

In the end, we will look closer to one more construction of the verb *chtít* (want). This construction mainly occurs in a spoken language and it is mentioned only briefly in grammar books, e.g. in (Karlík et al., 1995). Let us consider the following sentences:

(125) a. *Bábovka*$_{Pat}$ *se nechce péct.*
Cake$_{Nom}$ SE wants$_{Neg}$ bake$_{Inf}$.
'The cake refuses to get baked.'

b. *Bábovku$_{Pat}$ se mi nechce péct.*
Cake$_{Acc}$ SE me$_{Dat}$ wants$_{Neg}$ bake$_{Inf}$.
'I don't want to bake a cake'.

c. (*)*Bábovka se mi nechce péct.*
Cake$_{Nom}$ SE me$_{Dat}$ wants$_{Neg}$ bake$_{Inf}$.
'The cake refuses to get baked by me.' or
'I don't want to bake a cake.'

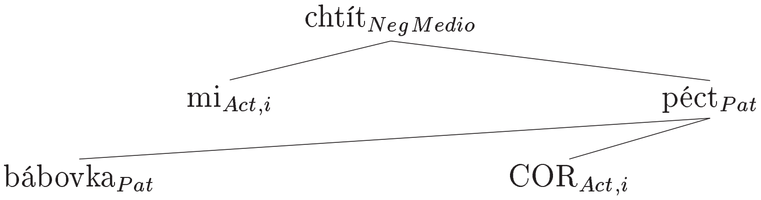
d. *Dort se mi nechce péct.*
Cake$_{Nom/Acc}$ SE me$_{Dat}$ wants$_{Neg}$ bake$_{Inf}$.

e. **Ten pán se mi nechce zdravit.*
That man$_{Nom}$ SE me$_{Dat}$ wants$_{Neg}$ greet$_{Inf}$.
'I don't want to greet that man.'

f. *Toho pána se mi nechce zdravit.*
That man$_{Acc}$ SE me$_{Dat}$ wants$_{Neg}$ greet$_{Inf}$.

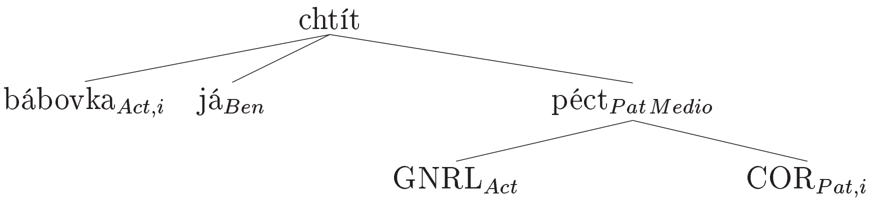
The sentence in (125a) is a construction with derived reflexive in infinitive. (125b) is very similar on the surface, but its syntactic structure is different. Here, the word *bábovka* (cake) is a direct object of the infinitive, and the whole infinitive clause is the subject of the verb *chtít*. The whole construction in the main clause is a mediopassive of the verb *chtít* with the Dative member as Actor. The structure is shown in (126):

(126)



The sentence in (125c) differs from the previous sentence by the case of the word *bábovka*, and from (125a) by the additional Dative member *mi* (me$_{Dat}$). We can understand the sentence as a variation of (125a), with Beneficiary expressed by Dative. The structure is shown in (127):

(127)



In colloquial speech, however, this construction is sometimes used in the meaning of (125b), although some speakers reject this construction. The problem with this sentence is that we have two candidates for the subject of the main clause. The first candidate is the word *bábovka*, which is in Nominative, and the second candidate is the infinitive *péct*, as in (125b). Our conclusion is that this construction is a result of misunderstanding of sentences like (125d), where the form of masculine inanimate noun *dort* is homonymous. The incorrectness of this construction is fully shown in (125e), where the position of the nominal subject is lexically occupied by a masculine animate noun. This sentence is out for all speakers.

We mentioned this construction here just as a peculiarity and we will not try to encode it in the lexicon.

5. Algorithm for processing the surface frames

In this chapter the automatic processing of the source data will be described. The format of the source data was described in Chapter 3. The desired content of the lexicon was described in Chapter 4. The steps which have to be done to achieve this are

1. identifying single frames
2. merging all variants of a single frame
3. marking the obligatority of frame members
4. assigning the functors to members
5. marking the possible diatheses

In the next sections these single steps will be described in detail.

5.1. Identifying and merging frames, marking the obligatority

In the source lexicon, every lemma is listed only once, even if it has several valency frames. A single valency frame, on the other hand, can have several variants (e.g. *učit koho co*$_{Acc}$, *učit koho čemu*$_{Dat}$—teach sb st). The variants of one frame are mixed with other frames and thus the first task is to separate the different frames and merge the variants. Let us show it with an example. The verb *bránit* has the following format in the source lexicon:

(128) **bránit** <v>hTc3,sI,hPc3-sUeN,hPc3-hTc6r{v}, (protect, prevent)
hPTc4,hPTc4-hPTc3r{proti},hPTc4-hPTc7r{před}

Now, we arrange the members of all its frames into a table (see Table 5.1): the rows are single "frames" from the original dictionary and the columns are single members of the frames. If there are more than one + in a column, then two or more frames share

	A hTc3	B sI	C hPc3	D sUeN	E hTc6r{v}	F hPTc4	G hPTc3r{proti}	H hPTc7r{před}
1	+							
2		+						
3			+	+				
4			+		+			
5						+		
6						+	+	
7						+		+

Table 5.1.: Identifying single frames

that member. Now, we try to find maximal non-intersecting parts. In Table 5.1 they are marked by the gray background. These gray parts represent real frames. Their members which never occur in one frame together can be declared with high probability as variants of one member (in Table 5.1) we can see that items D and E are variants of one member and items G and H are variants of another member). Now, we can merge the variants, which is shown in Table 5.2: the frames 3 and 4 were merged into 3′ and the frames 5 and 6 into 6′.

	A	B	C	D|E	F	G|H
1	+					
2		+				
3′			+	+		
5					+	
6′					+	+

Table 5.2.: Merging frame variants

There is a small problem with single-member frames (frames 1 and 2 in our example). They can be understood as separate frames, as in the case of *mířit kam* (head somewhere), *mířit na koho* (aim at sb), or as variants of one frame, as in the case of *bádat nad čím*, *bádat o čem* (research into st). We had to make a decision whether we wanted to merge all such frames, or whether we wanted to keep them separate. We decided to "merge as much as possible" because of an easier assignment of the functors, which will be explained in the next section. In our table, we then also merge the frames 1 and 2 into a frame with one member A|B.[1]

[1] A careful reader notices that the second frame should also contain Dative (hPc3) and it should in fact be merged with the third frame into one frame: **bránit [hPc3] [sI|sUeN]**. We showed here

In the above table we can also see how we identify obligatory members of a frame. In lines 5 and 6′, the member F is always present, while the other member G|H may be missing. Unfortunately, we are not able to say whether G|H is a general inner participant, or optional participant, or obligatory and deletable free modification, or even non-obligatory free modification, but at least the information about obligatory members of the frame should be correct. We use the square brackets for obligatory members of a frame (as was described in Chapter 4), and for now, we will use the parentheses for all other cases. The entry from example (128) now can be recorded as follows:

(129) a. `bránit [hTc3|sI]` (*bránit čemu/něco udělat*) (prevent st/doing st)

b. `bránit [hPc3] [sUeN]` (*bránit komu, aby něco neudělal*) (prevent sb from doing st)

c. `bránit [hPTc4] (hPTc3r{proti}|hPTc7r{před})` (*bránit koho/co {proti komu/čemu|před kým/čím}*) (protect sb/st {against sb/st|from sb/st})

As we said above, the source dictionary does not contain the so-called "left valency", i.e. subjects. This information is usually missing in printed dictionaries, as readers are able to fill the missing information, but in an electronic dictionary which is meant for language processing, such information must be included. We will describe the process of adding the subjects in the next section.

5.2. Assigning functors

It was shown by many authors that there is no straightforward correspondence between the deep frame and its surface realization. One can, however, try to find some regularities or tendencies, and then formulate rules for assigning the functors to the surface frames. The mappings between the tectogrammatical and morphemic levels (in active voice) is shown in Figure 5.1.

We can see that this picture does not help much—nearly everything is possible. It is necessary to add, however, that this picture also covers all marginal frames like líbit RSEs[i2]1(hPRc3)2[hPTc1]@ (like, appeal) and ubývat R--1[hTc2]@ (dwindle).[2]

Among all correspondences, there are some which are considered as typical. In the direction from the tectogrammatical level to the morphemic one these are Actor → Nominative, Patient → Accusative, Addressee → Dative, Effect → Instrumental, Origin →

a real example from the source lexicon, where some information was missing. The correction of this type of mistake is left for the post-editor.

[2] When we speak about marginal frames we do not say that the verbs with those frames are marginal, but the frames themselves are rather rare, and the lexicon contains only a few such frames. The verbs which have those frames may be in quite frequent use.

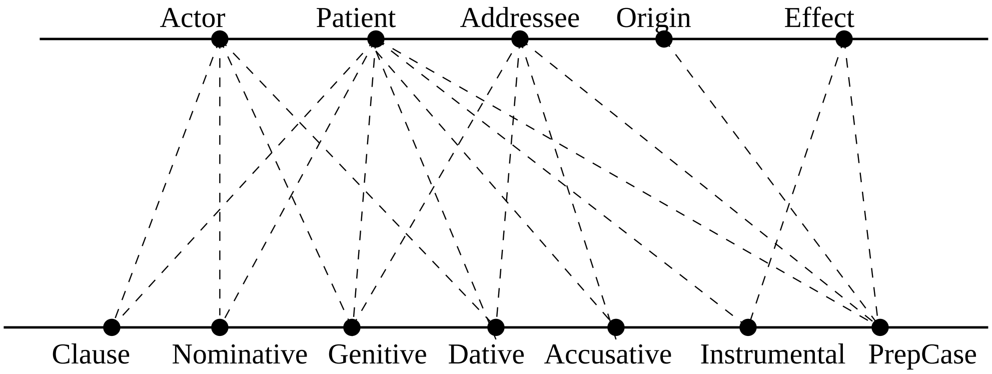


Figure 5.1.: Mapping between TL and ML in active voice

Genitive+Prep{z} (from) or Origin → Genitive+Prep{*od*} (from). In the opposite direction the correspondences are not so clear because of free modifications, which have a very broad repertory of the surface realizations. Thus Accusative can represent Patient or Temporal modification, Instrumental can represent Patient (*stát se*—become), Effect (*zvolit*—elect), Means (*zaplavit*—flood), Manner (*kopat*—dig); Genitive with the preposition *od* can represent Patient (e.g. *distancovat se*—dissociate), Origin (*dostat*—get), Direction from (*odejít*—leave), Temporal modification *how long* (*spát*–sleep), Cause (*opuchnout*—swell).

If we consider only frames with at least three actants[3] we get another picture shown in Figure 5.2.

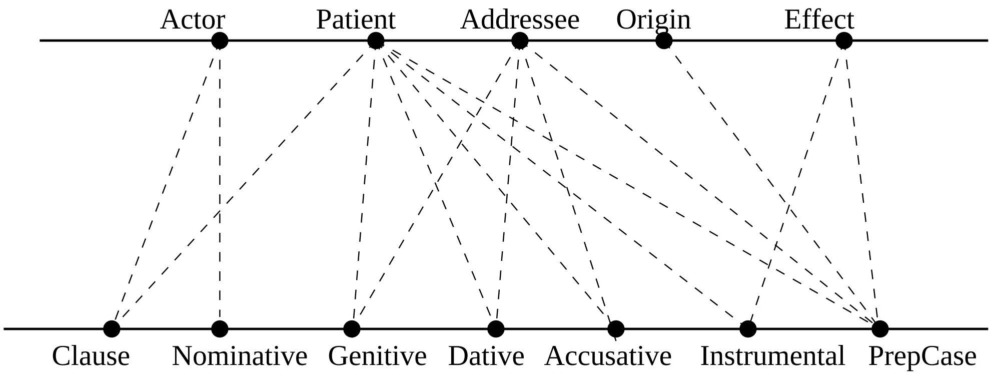


Figure 5.2.: Mapping between TL and ML for verbs with at least three actants

Though some joins disappeared, we still cannot find a unique mapping between the

[3]Frames with one or two actants are "uninteresting" as the functors are assigned after the rules listed in (6) in Chapter 2: if the frame has only one actant it is Actor, if there are two actants in the frame, they are Actor and Patient. In most cases, Actor is realized as Nominative and Patient as the "remaining" surface realization. There are some exceptional frames, as líbit RSEs[i2]1[hPRc3]2[hPTc1]@ (like, appeal) or zželet RSE1[hPc3]2[hPTRc2]@ (take pity on sb/st) which have to be edited manually.

	Actor	Patient	Addressee	Origin	Effect	
dát	(Nom)	Acc	Dat			give
dostat	Nom	Acc		<Gen+*od*>		get
šít	(Nom)	(Acc)	<Dat>	<Gen+*z*>		sew
předělat	(Nom)	Acc	<Dat>	<Gen+*z*>	<Acc+*na*>	remake
žádat	(Nom)	Acc		(Gen+*od*)		ask

Table 5.3.: Prototypical frames

tectogrammatical and morphemic level. However, we can observe that frames can be split in two groups. The first group contains verbs whose actants are only realized by typical surface forms; we call these frames *prototypical* (several examples are listed in Table 5.3). The other group contains verbs with *non-prototypical* frames, where at least one member is realized by a non-typical surface form (examples are in Table 5.4). This observation was done by J. Panevová, and an experimental algorithm for assigning the functors to surface realizations was created (see Panevová and Skoumalová, 1992). The algorithm checks whether a frame contains only prototypical surface forms, and if so it assigns them the corresponding functors. In Table 5.4, we can see that there is a possible source of problems in frames with surface forms in Accusative and Dative—their functors can be assigned the other way round than we expect. In this case we have to add one more criterion, and it is that Addressee must be "more animate" than Patient.[4] From this reason we only assume *animate* Dative as the typical realization of Addressee (hPc3 or hPTc3).

In the experiment, it was supposed that we worked only with inner participants (free modifications were filtered out), which made the task easier. In BRIEF lexicon, however, we cannot rely on getting actants only in surface frames, but on the other hand, the repertory of free modifications occurring in the lexicon is not as wide as in

[4] The scale of animacy (in BRIEF notation) is hT < hPT < hP.

	Actor	Patient	Addressee	Origin	Effect	
zvolit	(Nom)	Acc			Ins	elect
hrozit	(Nom)	Ins	(Dat)			threaten
vystavit	(Nom)	Dat	Acc			subject
dědit	(Nom)	(Acc)		(Loc+*po*)		inherit
hovořit	(Nom)	<Loc+*o*>	(Ins+*s*)			speak
psát	(Nom)	<Loc+*o*>	<Dat>		(Acc)	write
zeptat se	Nom	Acc+*na*	(Gen)			ask

Table 5.4.: Non-prototypical frames

the language as a whole (for example, a free modification of condition hardly occurs in a lexical entry). For this reason, we adopted a slightly different approach in the processing of BRIEF lexicon.

First, it was necessary to add the missing subjects. We did this automatically, and all frames got a subject in Nominative which was assigned the role of Actor: s[i1]1[hPTc1].[5]

The second step was assigning the roles to other members of the frame. Some preparation for this was done already while merging the frames: there is a list of possible functors for every surface realization, and this list was attached to every member of the original frame.[6] When we merged two members of a frame together we also made an intersection of the attached lists. An empty intersection prevented the two members from being merged. This process is shown in Table 5.5 on a frame of the verb *čertit se* (be angry). In BRIEF lexicon, the entry of this verb had the following form:

(130) čertit se <v>hPTc4r{na},hTc4r{pro},hTc7r{nad},hTc3r{kvůli}

	hPTc4r{na}	hTc4r{pro}	hTc7r{nad}	hTc3r{kvůli}
(ACTANT)	+	+	+	
DIR.WHERE	+			
CAUSE		+	+	+
PURPOSE			+	+
WHERE		+		

Table 5.5.: Merging frame of the verb *čertit se* (be angry)

Every surface realization is assigned a list of functors, as shown in the table. However, the functor ACTANT which denotes any actant is only taken in consideration if the surface realization has no variants.[7] As we first try to merge all the prepositional cases into one member of the frame, we exclude ACTANT from the list. In the rest of the table, we can see that the first prepositional case (hPTc4r{na}) has an empty intersection of functors with other prepositional cases which means that it cannot be taken as their

[5] Some Czech verbs do not have a subject at all, e.g. *pršet* (rain), in some frames the subject is realized by a clause or by an infinitive, e.g. *znamenat* (mean), *zdát se* (seem), but the vast majority of Czech verbs have a nominal subject in Nominative. The exceptions will be treated by a post-editor, again.

[6] These lists were created manually. The original lexicon was first divided into classes of frames containing a certain surface realization. These classes were analyzed and the surface realization was assigned a list of functors. Similar lists were also created for the Prague Dependency Treebank (Hajičová et al., 2000). These lists are longer because they contain all functors found in texts, not only in a lexicon. Beside it, they also contain more prepositional cases than the BRIEF lexicon.

[7] We do not try to assign single inner participants (Actor, Patient, etc.) in this step, we only mark whether a certain surface form can possibly represent an inner participants. Because of technical reasons we mark all potential inner participants as Patients—in a case that that there is only one actant beside Actor we get Patient "for free". In a case that there are more actants further processing is necessary.

variant inside one member of a frame. The remaining surface realizations have a non empty intersection of functors containing the value CAUSE. In the resulting frame, the first prepositional case will be assigned the functors ACTANT and DIR.WHERE. Other prepositional cases will be merged into one frame member which will be assigned the functor CAUSE:[8]

```
(131) čertit_se s[i1]1[hPTc1]2A[hPTc4r{na}] \
                 C[hTc4r{pro}|hTc7r{nad}|hTc3r{kvůli}]
```

After the merging of actants, we get three sorts of frames: frames where every member has only one functor assigned, frames where actants are distinguished from free modifications, but some of the free modifications are ambiguous, and frames where at least one member is ambiguous between an actant and a free modification. Approximately one third of all merged frames fall in the first category and another thousand into the second one. These frames are candidates for further processing with help of the above mentioned algorithm, and therefore they will be separated from the rest which must be left for post-editing.

Now, we will describe the process of assigning functors in the categories where actants are distinguished from free modifications. These frames fall into two subcategories: frames with at most two inner participants (i.e. Actor and Patient) and frames with at least three inner participants. The former are done already and we do not need to process them any further. The latter will be processed by the algorithm for assigning functors, but let us first resume the starting conditions:

- We have at least three inner participants.
- Actor is already assigned to the subject.
- We have to decide which of the actants is Patient and what are functors of the remaining inner participants.

We will not describe the algorithm in detail, we only sketch the overall strategy. More details and a flow chart can be found in Appendix D.

- A rule (following from the actant shifting) which must be observed after every step of the algorithm is that Patient slot must be filled. If there is only one unassigned member and the Patient slot has not been filled yet then the last member of the frame is assigned the Patient functor.
- We start with searching for Origin as Origin has the narrowest set of possible surface realizations, which in addition are not "polysemous".

[8]For the list of abbreviations used for functors see Appendix B.4, for lists of functors attached to every surface realization see Appendix C.2.

- Addressee assignment is ruled by the animacy of surface forms rather than the morphological cases. Animate Accusative or an animate prepositional case are realizations of Addressee rather than inanimate Dative.
- The decision about Effect can be quite difficult. Beside the typical prepositional cases also Instrumental can be a surface form of Effect. We then have to take into consideration the remaining unassigned members of the frame and make a decisions about pairs of surface forms.

As was said above, approximately 7500 frames are processed by this algorithm and the program ends successfully in all cases. The remaining ca 11,000 frames must be edited manually, with help of an editor prepared by Z. Žabokrtský (see Skoumalová et al., prep). The editor's work should be easier as s/he gets a (small) set of possible functors which can be assigned to every member of a frame and s/he does not have to choose from all 47 possibilities.

5.3. Marking diatheses

We made a simple assumption that

- reflexive verbs cannot have any diatheses (the exception with the periphrastic passive of the verb *tázat se* was discussed above), and so they get the mark `@`.
- intransitive verbs[9] can form reflexive passive; they get the mark `$`.
- a verb with a member in Accusative or in an indirect case (without preposition) can form both periphrastic and reflexive passive; it gets marks `%$`
- a verb whose all objects are prepositional cases can form the reflexive passive; it gets the mark `$`.

During the automatic processing all frames are assigned these marks and corrections will be made by the post-editor. Actors, which were added automatically to all frames, are marked as general ((`hPTc1`)) in frames that allow for forming any passive, and they are marked as obligatory ([`hPTc1`]) in other frames.

5.4. Usage of the final lexicon

The final product can be used in NLP systems for parsing, tagging, grammar checking and similar purposes. In all these applications, however, all possible instances of single frames must be generated. In the next section, it will be shown how we obtain single sentence patterns from frames.

[9] The term intransitive verb here means a verb with only one actant realized as subject in Nominative.

5.4.1. Generating frame instances from frames

The frames shown so far represent in fact sets of frame instances in which only one of all variants for every member is chosen, optional members may or may not be present, surface realizations of deletable or general participants can be missing and the passive constructions can be formed. Let us take some frames from (23), which we repeat here and let us add some more complicated cases:

(132) a. `akumulovat` `R--s[i1]1(hPTc1)2[hTc4]%$` (accumulate st)

b. `kazit~2` `RDEs[i1]1[hTc1]@` (decay)

c. `přihlásit~1` `R--s[i1]1(hPc1)2[hPTSRc4]A[hTc2r{do}|hTc4r{na}]%$` (enroll sb/st where)

d. `slíbit~1` `R--s[i1]1(hPc1)2[hZc4|sD|sIq3d%]3[hPc3]%$#*` (promise)

The frame in (132a) represents the following set of instances:[10]

(133) a. `R--s[i1]1[hPTc1]2[hTc4]`
Kamna akumulují teplo.
Radiator accumulates heat.

b. `P--s[i2]1[hPTc7]2[hTc1]`
Teplo je akumulováno kamny.
Heat is accumulated by radiator.

c. `P--s[i2]1[hG]2[hTc1]`
Teplo je akumulováno.
Heat is accumulated.

d. `PSEs[i2]1[hG]2[hTc1]`
Teplo SE akumuluje.
Heat SE accumulates.

The frame in (132b) has no possible variants, so the only operation will be the deletion of the diathesis mark:

(134) `RDEs[i1]1[hTc1]`
Potraviny se kazí.
Eatables SE decay.

The frame in (132c), on the other hand, represents quite a large set of frame instances:

[10] In the frame instances, we use only one type of brackets, as they only serve as delimiters.

(135) a. `přihlásit~1` `R--s[i1]1[hPc1]2[hPTSRc4]A[hTc2r{do}]`

b. `přihlásit~1` `R--s[i1]1[hPc1]2[hPTSRc4]A[hTc4r{na}]`

c. `přihlásit~1` `P--s[i2]1[hPc7]2[hPTc1]A[hTc2r{do}]`

d. `přihlásit~1` `P--s[i2]1[hPc7]2[hPTc1]A[hTc4r{na}]`

e. `přihlásit~1` `P--s[i2]1[hG]2[hPTc1]A[hTc2r{do}]`

f. `přihlásit~1` `P--s[i2]1[hG]2[hPTc1]A[hTc4r{na}]`

g. `přihlásit~1` `PSEs[i2]1[hG]2[hPTc1]A[hTc2r{do}]`

h. `přihlásit~1` `PSEs[i2]1[hG]2[hPTc1]A[hTc4r{na}]`

So far, we only needed marks that have been already defined, but for all instances of the verb *slíbit* (promise) we will also need marks for "frames" with the support verbs *mít* and *dostat*. For this purpose, three new marks for a type of a frame were introduced:

`M` — construction with the support verb *mít*

`D` — construction with the support verb *dostat*

`T` — resultative construction with the verb *mít*

Now, we can generate all instances of the frame:

(136) a. `slíbit~1` `R--s[i1]1[hPc1]2[sIq3d%]3[hPc3]`

b. `slíbit~1` `R--s[i1]1[hPc1]2[sD]3[hPc3]`

c. `slíbit~1` `R--s[i1]1[hPc1]2[hZc4]3[hPc3]`

d. `slíbit~1` `P--s[i2]1[hPc7]2[sIq3d%]3[hPc3]`

e. `slíbit~1` `P--s[i2]1[hPc7]2[sD]3[hPc3]`

f. `slíbit~1` `P--s[i2]1[hPc7]2[hZc1]3[hPc3]`

g. `slíbit~1` `P--s[i2]1[hG]2[sIq3d%]3[hPc3]`

h. `slíbit~1` `P--s[i2]1[hG]2[sD]3[hPc3]`

i. `slíbit~1` `P--s[i2]1[hG]2[hZc1]3[hPc3]`

j. `slíbit~1` `PSEs[i2]1[hG]2[sIq3d%]3[hPc3]`

k. `slíbit~1` `PSEs[i2]1[hG]2[sD]3[hPc3]`

```
l. slíbit~1     PSEs[i2]1[hG]2[hZc1]3[hPc3]
m. slíbit~1     M--s[i3]1[hPc2r{od}]2[sIq3d%]3[hPc3]
n. slíbit~1     M--s[i3]1[hPc2r{od}]2[sD]3[hPc3]
o. slíbit~1     M--s[i3]1[hPc2r{od}]2[hZc4]3[hPc3]
p. slíbit~1     M--s[i3]2[sIq3d%]3[hPc3]
q. slíbit~1     M--s[i3]2[sD]3[hPc3]
r. slíbit~1     M--s[i3]2[hZc4]3[hPc3]
s. slíbit~1     D--s[i3]1[hPc2r{od}]2[sIq3d%]3[hPc3]
t. slíbit~1     D--s[i3]1[hPc2r{od}]2[sD]3[hPc3]
u. slíbit~1     D--s[i3]1[hPc2r{od}]2[hZc4]3[hPc3]
v. slíbit~1     D--s[i3]2[sIq3d%]3[hPc3]
w. slíbit~1     D--s[i3]2[sD]3[hPc3]
x. slíbit~1     D--s[i3]2[hZc4]3[hPc3]
```

5.4.2. Extracting subcat lists

For testing whether our lexicon can be used also in other theoretical frameworks we made a small experiment with LFG. The verbs frames were converted to *templates* which can be used in a lexicon. These templates are then processed by lexical rules which derive all sentence patterns.

Every template contains a *predicate* (i.e. lemma and a subcat list) on which the lexical rules will be applied. A template can also contain some constraint which apply for all verbs of a given category. We will show it on an example:

(137)
```
TRANSRFLPERPASS(P) =
        @(LR-TRANSRFLPERPASS (^ PRED)='P<(^ SUBJ)(^ OBJ)>').
TRANSRFLPERPASSDAT(P) =
        @(LR-TRANSRFLPERPASS {(^ PRED)='P<(^ SUBJ)(^ OBJ)(^ OBJ2)>'
                              (^ OBJ2 CASE)=DAT}).
```

P in parentheses and in the subcat list is a variable for the lemma. The template TRANSRFLPERPASS is used for transitive verbs which have only one object and they can be passivized by both ways. The template TRANSRFLPERPASSDAT is used for transitive verbs which have another object in Dative and which can also be passivized by both ways. Both the templates use the same set of lexical rule, namely LR-TRANSRFLPERPASS:

(138)
```
LR-TRANSRFLPERPASS(SCHEMATA) =
        { SCHEMATA
        (^ OBJ CASE)=ACC
        ~(^ REFL)
        |SCHEMATA
        (^ REFL)=c SE
        (^ OBJ)->(^ SUBJ)
        (^ OBJ CASE)=NOM
        (^ SUBJ)->NULL
        |SCHEMATA
        (^ OBJ)->(^ SUBJ)
        (^ OBJ CASE)=NOM
        (^ SUBJ)->NULL
        ~(^ REFL)
        ~(^ TENSE)
        (^ PARTICIPLE)=c PASS }.
```

Lexical rules work like functions on variables supplied by templates. SCHEMATA stands for the variable and it is filled either by a predicate, or by a predicate an further constraints.

In the above example, we can see that three construction are created by the lexical rule LR-TRANSRFLPERPASS. The first construction is an active sentence where the object is in Accusative and the reflexive form of the verb is prohibited. The second construction is a reflexive passive, the object takes the position of a subject and the original subject is deleted. The third construction is a periphrastic passive, where, again, the object takes the position of a subject and the original subject is deleted, and further, the verb must have a form of passive participle and no reflexive particle can be part of the verb construction.

An experimental grammar was written for testing the lexicon. The lexicon only contains verbs from regular morphological paradigms so that the morphological module would not to be to large. Results of processing testing sentences are shown in Appendix F.

6. Conclusions

6.1. Verb grouping

One of the result of our work is categorization of Czech verbs according to their frames. Similar work was done by A. Horák (see Horák, 1998b) so we can compare our results.

Horák worked with the original BRIEF lexicon, and he did not try to separate single meanings of verbs or merge variants into one frame. In his approach, two verbs fall into one category if their lists of frames are identical. This means that every verb can only occur in one category, although it has several frames representing several meanings. Horák made four level of categorization: in 1st level, verbs are equivalent only if they share the same valency list; 2nd level has no surface realization of prepositional cases (they are supposed to be adjuncts and they are replaced by adverbial 'semantic' features). In 3rd level adjuncts are deleted from frames, and in 4th level the distinction between animate and inanimate members is removed.

	cl-brief		cl-opt		H 1st	
number of classes		3560		3978		4537
number of verbs		15022		15022		15022
number of valencies		NA		NA		49566
three biggest classes	⋆	(1435)	⋆⋆	(1435)	hTc4	(1420)
	†	(1000)	††	(1000)	hPTc4	(812)
	‡	(716)	‡‡	(685)	hTc7	(402)
no. of classes with 1 verb	1735	(48%)	1976	(49%)	2699	(59%)
no. of classes with 2 verbs	982	(27%)	1087	(27%)	1223	(27%)
no. of classes with 3 verbs	242	(6%)	276	(6%)	219	(5%)
no. of classes with more verbs	601	(19%)	639	(18%)	396	(9%)

⋆ R--s[i1]1[hPTc1]2[hTc4]%$ ⋆⋆ R--s[i1]1(hPTc1)2[hTc4]%$
† R--s[i1]1[hPTc1]2[hPTc4]%$ †† R--s[i1]1(hPTc1)2[hPTc4]%$
‡ R--s[i1]1[hPTc1]2[hPTc4]2CM[hTc7]%$ ‡‡ R--s[i1]1(hPTc1)2[hPTc4]2CM(hTc7)%$

Table 6.1.: Classification of verbs

We compare our results after merging the frames with Horák's 1st level classification. In Table 6.1 the column cl-brief describes verb frames which were merged but obligatority was not marked. The column cl-opt describes merged frames with obligatority, and the column H 1st shows Horák's results.

We can see that merging frames and identifying single meanings helps to decrease number of classes. It also changes the percentage of classes with one verb only and the number of classes with more verbs.

Next, we will work only with frames which were processed fully automatically and were fully resolved (this means every member of a frame is assigned only one inner participant or a free modification). Our set of verbs thus becomes smaller, but we can compare the percentage.

We removed all surface realizations of free modifications and left only the their functors in all frames. Such a lexicon can be compared with Horák's 2nd level of classification, and the comparison is done in Table 6.2.

	adjunct-brief		adjunct-opt		H 2nd	
number of classes		506		562		3188
number of verbs		6255		6255		15022
number of valencies		NA		NA		43175
three biggest classes	⋆	(1435)	⋆⋆	(1435)	hTc4	(1420)
	†	(1000)	††	(1000)	hPTc4	(812)
	‡	(586)	‡‡	(348)	hA	(553)
no. of classes with 1 verb	209	(41%)	236	(41%)	1780	(56%)
no. of classes with 2 verbs	117	(23%)	124	(22%)	884	(28%)
no. of classes with 3 verbs	45	(8%)	61	(10%)	155	(5%)
no. of classes with more verbs	135	(28%)	141	(27%)	369	(12%)

⋆ R–s[i1]1[hPTc1]2[hTc4]%$ ⋆⋆ R–s[i1]1(hPTc1)2[hTc4]%$
† R–s[i1]1[hPTc1]2[hPTc4]%$ †† R–s[i1]1(hPTc1)2[hPTc4]%$
‡ R–s[i1]1[hPTc1]2[hTc4]3[hPc3]%$ ‡‡ R–s[i1]1(hPTc1)2[hTc4]3[hPc3]%$

Table 6.2.: Classification of verbs with adjuncts simplified

We can see that the number of classes with one verb only decreased again. The next step in Horák's work was to delete the hypothetical free modifications, but we do not think that this improves the classification. The number of classes decreases, but to the detriment of the accuracy. As free modifications can be obligatory participants of verb frames we cannot just delete all of them.

The next Horák's step is suppressing the difference between animate and inanimate participants. We again do not agree that this helps to improve the results of sorting. Let us consider the verb *vystavit*:

(139) a. *vystavit co komu*
issue st_{Acc} sb_{Dat}

b. *vystavit koho čemu*
subject sb_{Acc} st_{Dat}

The surface realizations of the two constructions differ only in animacy of the participant, so we should not to get rid of this information. We believe that the way which leads to better frame categorization is in improving the description of the verbs.

6.2. Further perspectives

We have shown a syntactic lexicon which can be used in various systems of natural language processing, especially in systems using symbolic methods (as opposed to stochastic methods). The lexicon, however, still needs some editing work, but we believe that it was pre-processed in such a way that the editing work will be easy. In the near future, the following things will be done:

1. An editor will be created with the help of which it will be possible to add new verbs and frames and to correct those already stored in the lexicon. A prototype of the editor already exists and now it is being tested on a small sample of the lexicon.
2. Some verb classes are categorized now as ambiguous, but they are homogeneous in that sense that the ambiguity can be resolved the same way for the whole class. This is the case of, for example, the class containing verbs with the frame R--s[i1]1(hPTc1)2[hTc4]2CM(hTc7)%$—the ambiguous member hTc7 will be assign the functor Means for the whole class. It will be necessary to go through all classes and decide which of them can be resolved this way.
3. So far, only two types of diatheses have been added to verb frames. We also have to enhance the lexical entries by the information on other types of diatheses.
4. We have left aside idioms and phrases, but we have to include them in the lexicon as well. One problem is that they are not encoded in a unique way, and another that we have to create a format for them. There are also some theoretical problems with idioms, as whether the expressions in the "fixed" part of an idiom should be assigned functors or not, etc.

The lexicon can be used in many NLP applications, as parsing or tagging of Czech texts. One of the challenges is tagging of the Czech National Corpus. It has been tagged with the help of stochastic methods, but the results of the tagging made several linguists start work on a rule-based tagger (see Oliva et al., 2000). Though their achievements are admirable, a syntactic lexicon can improve their results even more.

Bibliography

Avgustinova, T., Bémová, A., Hajičová, E., Oliva, K., Panevová, J., Petkevič, V., Sgall, P., and Skoumalová, H. (1995). Linguistic problems of Czech. Technical report, JRP PECO 2824, Prague. Final research report.

Bémová, A. (1984). Verbal prefixation from the viewpoint of valency. In Sgall (1984), pages 275–287.

Benson, M., Benson, E., and Ilson, R. (1986). *The BBI Combinatory Dictionary of English*. John Benjamins Publishing Company, Amsterdam/Philadelphia.

Borsley, R. (1999). *Syntactic Theory: A Unified Approach*. Arnold, London, 2 edition.

Čermák, F. (1997). *Jazyk a jazykověda (Language and Linguistics)*. Pražská imaginace, Prague.

Čermák, F. and Holub, J. (1991). *Syntagmatika a padigmatika českého slovesa I. Valence a kolokabilita (Syntagmatics and Paradigmatics of Czech Verb I. Valency and collocability)*. Charles University, Faculty of Philosophy, Prague.

Chomsky, N. (1986). *Knowledge of Language: Its Nature, Origin, and Use*. Convergence. Praeger, Westport.

Comrie, B. (1991). On so-called raising in Russian. In Grochowski, M. and Weiss, D., editors, *Words are Physicians for Ailing Mind (= Sagners Slavistische Sammlung 17)*, pages 121–128. O. Sagner, München.

Dalrymple, M., Kaplan, R. M., Maxwell III, J. T., and Zaenen, A., editors (1995). *Formal Issues in Lexical-Functional Grammar*. Number 47 in Lecture Notes. CSLI, Stanford.

Daneš, F., Hlavsa, Z., et al. (1987a). *Větné vzorce v češtině (Sentential paradigms in Czech)*. Number 23 in Studie a práce lingvistické. Academia, Prague.

Daneš, F., Hlavsa, Z., Grepl, M., et al. (1987b). Mluvnice češtiny 3—Skladba (Grammar of Czech 3—Syntax). Academia, Prague.

Dokulil, M. (1941). Morfologické kategorie pasiva ve spisovných jazycích severských ve srovnání se spisovnou češtinou (Morphological categories of passive in Nordic standard languages in comparison with standard Czech). In *Hrst studií a vzpomínek: prof. dr. Ant. Beerovi jeho žáci (Handful of studies and memories: to prof. dr. A. Beer from his pupils)*, pages 77–99. Odbočka Jednoty českých filologů v Brně, Brno.

Fillmore, C. J. (1968). The case for case. In Bach, E. and Harms, R. T., editors, *Universals in Linguistic Theory*. Holt, Rinehart and Winston.

Grepl, M. and Karlík, P. (1989). *Skladba spisovné češtiny (Syntax of Standard Czech)*. SPN, Prague, 2 edition.

Grepl, M. and Karlík, P. (1998). *Skladba češtiny (Syntax of Czech)*. Votobia, Olomouc.

Hajič, J. (1994). *Unification Morphology Grammar*. PhD thesis, Charles University, Faculty of Mathematics and Physics, Prague.

Hajič, J., Panevová, J., Buráňová, E., Urešová, Z., Bémová, A., Štěpánek, J., Pajas, P., and Kárník, J. (1999). *Anotace na analytické rovině: Návod pro anotátory (Annotation on analytical level: Manual for the annotators)*. Charles University, Faculty of Mathematics and Physics, Prague. URL: http://ufal.mff.cuni.cz/projekty.html.

Hajičová, E. and Panevová, J. (1984). Valency (case) frames of verbs. In Sgall (1984), pages 147–188.

Hajičová, E., Panevová, J., and Sgall, P. (2000). A Manual for Tectogrammatical Tagging of the Prague Dependency Treebank. Technical Report ÚFAL/CKL TR-2000-09, Charles University, Faculty of Mathematics and Physics, Prague. URL: http://ufal.mff.cuni.cz/pdt/pdt.html.

Hausenblas, K. (1963). Slovesná kategorie výsledného stavu v dnešní češtině (verb category of resultative in contemporary Czech). *Naše řeč*, 46:13–28.

Havránek, B. (1928). *Genera verbi v jazycích slovanských I (Genera verbi in Slavic languages I*. Královská česká společnost nauk, Prague.

Horák, A. (1998a). Popis formátu brief (Description of the format of the lexicon). Unpublished documentation.

Horák, A. (1998b). Verb valency and semantic classification of verbs. In Sojka, P., Matoušek, V., Pala, K., and Kopeček, I., editors, *Proceedings of the First Workshop on Text, Speech, Dialogue — TSD'98*, pages 61–66. Masaryk University Press, Brno.

Kaplan, R. M. and Maxwell, J. T. (1996). LFG grammar writer's workbench. Technical report, Xerox PARC, Palo Alto.

Karlík, P. (2000). Hypotéza modifikované valenční teorie (Hypothesis of the modified valency theory). *Slovo a slovesnost*, LXI(3):170–189.

Karlík, P., Nekula, M., and Rusínová, Z., editors (1995). *Příruční mluvnice češtiny (Handbook of Czech Grammar).* Nakladatelství Lidové Noviny, Prague.

Kathol, A. (1994). Passives without Lexical Rules. In Nerbonne et al. (1994), pages 237–272.

Kocek, J., Kopřivová, M., and Kučera, K., editors (2000). *Český národní korpus — Úvod a příručka uživatele.* Charles University, Faculty of Philosophy, Prague. URL: http://ucnk.ff.cuni.cz.

Králíková, K. (1980). Pasívum v generativním popisu češtiny (Passive voice in the generative description of Czech). *Slovo a slovesnost*, XLI(2):118–126. Translated in Sgall (1984).

Králíková, K. (1981). Reflexívnost sloves z hlediska automatické analýzy češtiny (Reflexivity of verbs from the point ofperspective of automatic analysis of Czech). *Slovo a slovesnost*, XLII(4):291–298.

Kuboň, V. (1999). A robust parser for Czech. Technical Report ÚFAL TR-1999-06, Charles University, Faculty of Mathematics and Physics, Prague.

Luelsdorff, P. A., Panevová, J., and Sgall, P., editors (1994). *Praguiana 1945-1990.* Number 40 in Linguistic an Literary Studies in Eastern Europe. John Benjamins Publishing Company, Amsterdam/Philadelphia.

Nerbonne, J., Netter, K., and Pollard, C., editors (1994). *German in Head-Driven Phrase Structure Grammar.* Number 46 in Lecture Notes. CSLI, Stanford.

Oliva, K. (1989). *A Parser for Czech Implemented in Systems Q.* Number XVI in Explizite Beschreibung der Sprache und automatische Textbearbeitung. Charles University, Faculty of Mathematics and Physics, Prague.

Oliva, K. (1994). HPSG Lexicon without Lexical Rules. In *Proceedings of COLING '94*, pages 823–826, Kyoto.

Oliva, K. (1996). A grammar checker for Czech. Technical report, JRP PECO 2824, Prague.

Oliva, K. (2000). Hovory k sobě/si/sebe/se (Discussion on sobě/si/sebe/se). In Hladká, Z. and Karlík, P., editors, *Čeština—univerzália a specifika 2, (Czech—Universals and Specifics 2)*, Proceedings of the Conference held in Šlapanice u Brna, November 17-19, 1999, pages 167–171, Brno. Masaryk University.

Oliva, K., Hnátková, M., Petkevič, V., and Květoň, P. (2000). The linguistic basis of a rule-based tagger of Czech. In Sojka, P., Kopeček, I., and Pala, K., editors, *Proceedings on the Third International Workshop on Text, Speech and Dialogue—TSD 2000*, Lecture Notes in Artificial Intelligence LNCS/LNAI 1902, pages 3–8. Springer-Verlag.

Pala, K. and Ševeček, P. (1997). Valence českých sloves (Valency of Czech verbs). In *Sborník prací FFBU*, volume A45, pages 41–54. Masaryk University, Brno.

Panevová, J. (1971). Časové a vidové kategorie predikátu (Tense and aspect categories of predicate). In Panevová et al. (1971), pages 23–44.

Panevová, J. (1974-75). On verbal frames in functional generative description, Part I and II. *Prague Bulletin of Mathematical Linguistics*, 22:3-40,23:17-52.

Panevová, J. (1980). *Formy a funkce ve stavbě české věty (Forms and Functions in Syntax of Czech Sentence).* Number 13 in Studie a práce lingvistické. Academia, Prague.

Panevová, J. (1984). Sentences with general actor. In Sgall (1984), pages 203–221.

Panevová, J. (1996). More remarks on control. In Hajičová, E., Leška, O., Sgall, P., and Skoumalová, Z., editors, *Prague Linguistic Circle Papers*, volume 2, pages 101–120. John Benjamins Publishing Company, Amsterdam/Philadelphia.

Panevová, J. (1998). Ještě k teorii valence (Once more to the theory of valency). *Slovo a slovesnost*, LIX(1):1–14.

Panevová, J. (1999). Česká recipročni zájmena a slovesná valence (Czech reciprocal pronouns and verb valency). *Slovo a Slovesnost*, LX(4):269–275.

Panevová, J., Benešová, E., and Sgall, P. (1971). *Čas a modalita v češtině (Tense and Modality in Czech).* Universita Karlova, Prague.

Panevová, J. and Skoumalová, H. (1992). Surface and deep cases. In *Proceedings of COLING '92*, pages 885–889, Nantes.

Pauliny, E. (1943). *Štruktúra slovenského slovesa (Structure of the Slovak verb).* Slovenská akadémia vied a umení, Bratislava. Translated and in Luelsdorff et al. (1994).

Pollard, C. and Sag, I. A. (1987). *Information-Based Syntax and Semantics, Volume 1, Fundamentals.* Number 13 in Lecture Notes. CSLI, Stanford.

Pollard, C. and Sag, I. A. (1994). *Head-Driven Phrase Structure Grammar.* The University of Chicago Press, Chicago & London.

Sgall, P. (1967). *Generativní popis jazyka a česká deklinace (Generative Description of Language and Czech Declension).* Number 6 in Studie a práce lingvistické. Československá akademie věd.

Sgall, P., editor (1984). *Contributions to Functional Syntax, Semantics, and Language Comprehension*, volume 16 of *Linguistic & Literary Studies in Eastern Europe.* John Benjamins Publishing Company, Amsterdam/Philadelphia.

Sgall, P. (1998). Teorie valence a její formální zpracování (Theory of valency and its formalization). *Slovo a slovesnost*, LIX(1):15–29.

Sgall, P., Hajičová, E., and Panevová, J. (1986). *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects.* D. Reidel Publishing Company, Dordrecht.

Skoumalová, H. (1994). Czech dictionary for the grammar checker. Technical report, JRP PECO 2824, Saarbrücken.

Skoumalová, H. (1997). Verb frames in the Czech hierarchical lexicon. *TELRI Newsletter*, 6:18–32.

Skoumalová, H. (1998). Derived frames and the lexicon. In Hajičová, E., editor, *Issues of Valency and Meaning—Studies in Honour of Jarmila Panevová*, pages 154–168. Karolinum, Prague.

Skoumalová, H., Straňáková-Lopatková, M., and Žabokrtský, Z. (in prep.). Enhancig the valency dictionary of Czech verbs: Tectogrammatical annotation.

Šmilauer, V. (1967). *Novočeská skladba (Syntax of Modern Czech).* Academia, Prague.

SSJČ (1989). *Slovník spisovného jazyka českého (Dictionary of standard Czech).* Academia, Prague.

Štícha, F. (1984). *Utváření a hierarchizace struktury větného znaku (Creation and hierarchization of the structure of a sentence sign).* Univerzita Karlova, Prague.

Straňáková-Lopatková, M. (2001). *Homonymie předložkových skupin a možnost jejího automatického zpracování (Homonymy of prepositional groups and possibility of its automatic processing).* PhD thesis, Charles University, Faculty of Mathematics and Physics, Prague.

Svoboda, K. (1962). *Infinitiv v současné spisovné češtině (Infinitive in Contemporary Standard Czech).* Rozpravy ČSAV. Academia, Prague.

Svozilová, N., Prouzová, H., and Jirsová, A. (1997). *Slovesa pro praxi (Verbs for practical use).* Academia, Prague.

Tesnière, L. (1959). *Eléments de syntaxe structurale.* Klincksieck, Paris.

Vidugyrite, V. (1999). Reflexivní slovesa: Srovnání češtiny s litevštinou (Reflexive verbs: Comparison of Czech and Lithuanian). Master's thesis, Charles University, Faculty of Philosophy, Prague.

Žabokrtský, Z. (2000). Dva experimenty s valenčním slovníkem (Two experiments with a valency dictionary). Unpublished article.

Subject index

Verbs used in examples

A. Abbreviations

1Pl — 1^{st} person, plural

1Sg — 1^{st} person, singular

2Pl — 2^{nd} person, plural

2Sg — 2^{nd} person, singular

3Pl — 3^{rd} person, plural

3Sg — 3^{rd} person, singular

Acc — Accusative

Act — Actor

Addr — Addressee

Adj — adjective

AVM — attribute-value matrix

Ben — Beneficiary

CNC — Czech National Corpus

COORD — coordination

COR, Cor — coreference, coreferential participant

Dat — Dative

Eff — Effect

Fem — feminine

FGD — Functional Generative Description

Fut — future tense

GB — Government-Binding Theory

GNRL, Gnrl — general participant

Gen — Genitive

Imp — imperative

imperf. — imperfective aspect

Inf — infinitive

Ins — Instrumental

LFG — Lexical-Functional Grammar

Loc — Locative

Masc — masculine

ML — morphemic level

Medio — mediopassive

Neg — negation

Neut — neuter

NLP — natural language processing

Nom — Nominative

Orig — Origin

Pat — Patient

PDT — Prague Dependency Treebank

perf. — perfective aspect

Pl — plural

Prtcpl — (passive) participle

RECP, Recp — reciprocity, reciprocal

Sg — singular

TL — tectogrammatical level

B. Symbols used in the dictionary

B.1. Voice

`R` — regular frame (in active voice with possible derivations)

`P` — irregular passive frame

There are three more marks which are not used in the lexicon, but they are exploited in sentence patterns generated from the frames.

`M` — construction with support verb *mít*

`D` — construction with support verb *dostat*

`T` — resultative construction with verb *mít*

B.2. Reflexivity

`--` — no reflexive particle; no reciprocity

`SE` — reflexive tantum with particle *se* (*bát se*)

`DE` — derived reflexive with particle *se* (*vlévat se*)

`se` — reflexive with optional particle *se* (*koukat se*)

`SI` — reflexive tantum with particle *si* (*stěžovat si*)

`DI` — derived reflexive with particle *si* (*vynachválit si*)

`si` — reflexive with optional particle *si* (*myslet si*)

B.3. Subject

`s` — subject; the attribute in brackets shows the type of the subject and its value points to functor which is cirrently the subject

- `i` — inherent
- `a` — raised

B.4. Functors

`1` — Actor

`2` — Patient

`3` — Addressee

`4` — Origin

`5` — Effect

`0` — no functor; used in frames of raising verbs

`A` — direction where (kAm)

`B` — Beneficiary

`C` — Cause

`D` — how long (jakDlouho)

`E` — where (kdE)

`F` — diFference

`G` — reGard

`H` — Heritage

`I` — Intent

`J` — how (Jak)

`K` — *reserved*

`L` — *reserved*

`M` — Means

`N` — Norm

`O` — from where (Odkud)

`P` — intent (Purpose, aim)

`Q` — *reserved*

`R` — compaRison

`S` — Substitution

`T` — criTerion

`U` — which way (kUdy)

`V` — accompaniment (průVod)

`W` — *reserved*

`X` — eXtent

`Y` — when (kdY)

`Z` — from when (Zekdy)

B.5. Grammatemes

`h` — ‘semantic’ features

- `P` — person
- `T` — thing, animal
- `S` — short reflexive pronoun *se* or *si*
- `R` — long reflexive pronoun *sebe*, *sobě*, etc.
- `Z` — interrogative pronoun *co* (what), demonstrative pronoun *to* (that), *všechno* (everything), etc.
- `G` — general participant (used in irregular passive frames and in generated sentence patterns)
- `E` — deleted (empty, erased) participant (used in generated sentence patterns)
- `C` — direct speech
- `Q` — quality (adjective)
- `M` — quantity (number, figure)
- `L` — location (adverb)

- `A` — direction where (adverb)
- `F` — direction from where (adverb)
- `D` — which way (adverb)
- `W` — when (adverb)

`c` — case

- `1` — Nominative
- `2` — Genitive
- `3` — Dative
- `4` — Accusative
- `6` — Locative
- `7` — Instrumental

`r` — preposition

`n` — number

- `S` — singular
- `P` — plural

`s` — clause

- `I` — infinitive
- `C` — conjunction *až*
- `D` — conjunction *že*
- `F` — conjunction *jestli*, *zda*
- `P` — conjunction *ať*
- `R` — relative expression *co*, *který*, *kdo*, ...
- `U` — conjunction *aby*
- `Z` — conjunction *jak*

`l` — required lemma

`e` — negation of a clause

- `A` — affirmative (default)
- `N` — negative

`x` — reciprocal coreference; the value points to a coindexed functor

`a` — subject raised to object position; the value points to the embeded clause from which the subject was raised `q` — subject- or object-control

`p` — “patient” control

`t` — “addressee” control

`d` — diatheses of embeded infinitive; the values are identical with values of the “main” frame

`m` — modality

- `D` — debitive (*muset*)
- `H` — hortative (*mít*)
- `V` — volitive (*chtít*)
- `P` — possibilitive (*moci*)
- `R` — permissive (*smět*)
- `F` — facultative (*dovést*)

B.6. Obligatority

`[ ]` — obligatory participant

`( )` — obligatory inner participant which can be realized as general, or obligatory and deletable free modification

`< >` — optional participant

B.7. Passive and other diathesis

`%` — periphrastic passive is possible (*číst*, *stavět*)

`$` — reflexive passive is possible (*číst*, *mluvit*, *jít*)

`@` — no passive (*bát se*)

`#` — constructions with *mít* (*slíbit*)

`*` — constructions with *dostat* (*vynadat*)

`~` — constructions with resultative *mít* (*uvařit*)

C. Possible functors assigned to grammatemes

C.1. Abbreviations used in lists of possible functors

`X` — Unknown functor; mostly error in source data.

`PAT` — Any actant. The reason why we chose this abbreviation is purely technical and it was explained in footnote 7 in Chapter 5.

`KAM` — Direction 'to'.

`BEN` — Beneficiary.

`CAUSE` — Cause.

`JAKDL` — Temporal modification 'how long'.

`KDE` — Location 'where'.

`DIFF` — Difference.

`REGARD` — Regard.

`HER` — Heritage.

`INT` — Intent.

`JAK` — Manner.

`MEANS` — Means.

`NORM` — Norm.

`ODKUD` — Direction 'from'.

`PURP` — Purpose.

`COMPAR` — Comparison.

`SUBST` — Substitution.

`CRIT` — Criterion.

`KUDY` — Direction 'which way'.

`ACCOMP` — Accompaniment.

`EXTENT` — Extent.

`KDY` — Temporal modification 'when'.

`ZEKDY` — Temporal modification 'from when'.

C.2. Lists of functors attached to every surface realization

Functors in parentheses are only taken in consideration if the surface realization has no variants. For example the prepositional case Accusative+*na* is typically a surface realization of direction, but in the frame of the verb *spoléhat na koho/co* (rely on sb/st) it is Patient.

The order of surface realization is important. A realization which is higher is listed first in brackets with variants and it are taken as a "representant" of the whole frame member.

```
hPc2    PAT
hPTc2   PAT
hTc2    PAT
v{eN}hTc2       PAT
hPc4    PAT
hPTc4   PAT
hTc4    PAT
sD      PAT
sF      PAT
sP      PAT
sPeN    PAT
sR      PAT
sUeN    PAT
sZ      PAT
sI      PAT INTENT KAM
sU      PAT PURP
sC      JAKDL
hA      KAM
hF      ODKUD
hL      KDE
hM      PAT
hPc1    PAT
hPTc1   PAT
hQc1    PAT
hQc7    PAT
hPc3    PAT
hPTc3   PAT
hTc3    (PAT) PURP
hPc7    PAT JAK
hPTc7   (PAT) MEANS SUBST
hRc7    PAT
hTc7    (PAT) MEANS CAUSE
hMr{na} PAT
hMr{o}  DIFF
```

```
hMr{za} MEANS
hAr{do} KAM
hAr{na} KAM
hPc3r{vůči}     PAT
hPc4r{o}        PAT
hPc6r{o}        PAT
hPc6r{po}       PAT
hPc6r{při}      PAT
hPc6r{v}        KDE
hPc7r{mezi}     (PAT) MEANS KDE
hPc7r{za}       KAM
hPTc1r{jako}    JAK
hPTc2r{bez}     (PAT) JAK
hPTc2r{do}      (PAT) KAM
hPTc2r{místo}   SUBST
hPc2r{u}        KDE
hPTc2r{u}       (PAT) KDE
hPTc2r{vedle}   KAM
hPc2r{kolem}    KDE
hPc3r{proti}    PAT KAM BEN
hPc7r{nad}      (PAT) KDE CAUSE
hPc7r{pod}      KDE
hPc7r{před}     (PAT) KDE
hPTc2r{kolem}   KDE KUDY
hPTc2r{od}      (PAT) ODKUD
hPc2r{od}       (PAT) ODKUD
hPc2r{z}        PAT
hPTc2r{z}       (PAT) ODKUD
hPc6r{na}       PAT
hPc3r{ke}       (PAT) KAM
hPc2r{do}       PAT
hPc4r{mezi}     (PAT) KAM
hPc4r{nad}      KAM
hPc4r{na}       PAT BEN
hPc4r{před}     PAT
hPTc3r{ke}      (PAT) KAM
hPTc3r{k}       (PAT) KAM
hPc4r{za}       PAT SUBST
hPc4r{pro}      PAT BEN
hPc7r{s}        (PAT) ACCOMP
hPTc3r{kvůli}   CAUSE
hPTc3r{proti}   PAT
hPTc4r{jako}    (PAT) JAK
hPTc4r{mezi}    KAM
hPTc4r{nad}     (PAT) KAM KDE JAK
hPTc4r{na}      (PAT) KAM
```

hPTc4r{o}	PAT
hPTc4r{pod}	KAM
hPTc4r{pro}	PAT BEN
hPTc4r{před}	KAM
hPTc4r{přes}	KAM KUDY
hPTc4r{v}	PAT
hPTc4r{za}	(PAT) SUBST
hPTc6r{na}	(PAT) KDE
hPTc6r{o}	PAT
hPTc6r{po}	(PAT) HER KAM
hPTc6r{při}	PAT
hPTc6r{v}	(PAT) KDE
hPTc7r{mezi}	(PAT) KDE KUDY
hPTc7r{nad}	(PAT) KDE KUDY CAUSE
hPTc7r{pod}	KDE
hPTc7r{před}	(PAT) PURP
hPTc7r{s}	(PAT) ACCOMP
hPTc7r{za}	(PAT) KDE KAM
hRc2r{od}	ODKUD
hRc2r{ze}	PAT
hRc3r{k}	KAM
hRc4r{pod}	KAM
hRc4r{pro}	JAK
hRc4r{ze}	ODKUD
hRc7r{mezi}	PAT
hRc7r{před}	JAK
hRc7r{s}	JAK
hTc2r{bez}	JAK ACCOMP
hTc2r{během}	KDY
hTc2r{do}	(PAT) KAM
hTc2r{kolem}	KDE KUDY JAK
hTc2r{od}	(PAT) ODKUD JAKDL ZEKDY CAUSE
hTc2r{podle}	NORM CRIT
hTc2r{podél}	KDE KUDY
hTc2r{pomocí}	MEANS
hTc2r{u}	KDE
hTc2r{vedle}	KDE ACCOMP
hTc2r{z}	(PAT) ODKUD
hTc3r{kvůli}	CAUSE PURP
hTc3r{k}	PAT KAM PURP
hTc3r{proti}	PURP BEN
hTc3r{vzhledem k}	REGARD
hTc4r{jako}	(PAT) COMPAR JAK
hTc4r{mezi}	KAM
hTc4r{mimo}	KDE KAM
hTc4r{nad}	KAM

```
hTc4r{na}       (PAT) KAM PURP
hTc4r{o}        (PAT) KAM DIFF
hTc4r{pod}      KAM
hTc4r{po}       EXTENT JAKDL
hTc4r{pro}      (PAT) PURP CAUSE
hTc4r{před}     KAM
hTc4r{přes}     KAM KUDY MEANS JAK
hTc4r{skrze}    KUDY
hTc4r{skrz}     KUDY
hTc4r{v}        (PAT) KAM
hTc4r{za}       (PAT) KAM JAK CAUSE
hTc6r{jako v}   JAK
hTc6r{na}       (PAT) KDE JAK
hTc6r{o}        PAT KDY JAK
hTc6r{po}       KDY KAM KUDY JAK CRIT
hTc6r{při}      KDE KDY
hTc6r{v}        (PAT) ACCOMP KDE JAK
hTc7r{mezi}     (PAT) KUDY KDE
hTc7r{nad}      (PAT) KDE CAUSE
hTc7r{pod}      ACCOMP KDE KUDY CAUSE
hTc7r{před}     (PAT) KDY KDE KUDY
hTc7r{s}        (PAT) MEANS ACCOMP
hTc7r{za}       (PAT) KDE KAM
v{eN}hPTc4r{na} PAT
v{eN}hTc2r{do}  KAM
v{eN}hTc3r{k}   PAT
hTc6    X
hPc3r{o}        X
hTc2r{v}        X
hTc7r{v}        X
hPTc4r{do}      X
hRc4r{do}       X
hRc4r{kolem}    X
hTc3r{v}        X
hTc4r{a}        X
```

D. Algorithm for assigning functors

D.1. Prototypical and less typical surface forms

Every participant which is in a typical form is assigned the corresponding functor (i.e. Patient, Addressee, Origin and Effect, resp.). If all participants were assigned a functor and (one) Patient is among them we have a prototypical frame, and the frame is done.

Typical forms:

Patient: `hPc2`, `hPTc2`, `hTc2`, `veNhTc2`, `hPTc4`, `hTc4`, `sD`, `sF`, `sP`, `sPeN`, `sR`, `sUeN`, `sZ`, `hM`, `hPc1`, `hPTc1`, `hQc1`, `hQc7`

Addressee: `hPc3`, `hPTc3`

Origin: `hPc2r{z}`, `hPTc2r{z}`, `hTc2r{z}`, `hPc2r{od}`, `hPTc2r{od}`, `hTc2r{od}`

Effect: `hTc4r{na}`, `hPTc4r{na}`, `hTc4r{v}`, `hPTc4r{v}`

If there are members with no assignment we check whether they are among "less typical" surface forms. Then we check again whether all participants were assigned a functor and whether Patient is among them.

Less typical forms:

Patient: `sI`, `sU`, `hPc4r{na}`, `hPc4`, `hPc6r{o}`, `hPTc6r{o}`, `hTc6r{o}`

Addressee: `hPc3r{proti}`, `hPTc3r{proti}`

Origin: `hPc6r{na}`, `hPc6r{po}`

Effect: `hTc4r{o}`, `hPTc4r{o}, hPc4r{o}`

The reason why we look for prototypical frames in two steps is simple. If we fail to assign the prototypical frame we have to continue with a non-protoypical frame. During the assignment of the non-protoyypical frame the prototypical forms take precedence over the less typical forms.

D.2. Assigning non-prototypical frame

In Figure D.1 we can see the algorithm for assigning actants to non-prototypical frame, if the frame contains at most one form typical for Patient.

If two members have forms typical for Patient special treatment is needed. In such a case we have to decide whether the two members are Addressee and Patient (e.g. *přesvědčit koho*$_{Addr}$ *aby/ať*$_{Pat}$—persuade sb to do st) or Patient and Effect (e.g. *dozvědět se co*$_{Eff}$*na koho/o kom*$_{Pat}$—to learn st about sb). The main criterion is animacy. If one of the members has the form hPc4 or hPTc4 it is declared Addressee. In other cases we decide between Patient and Effect: the "more animate" member is declared Patient and the other Effect.

D.3. Results

A small lexicon containing 105 most frequent verbs from CNC (Kocek et al., 2000) was extracted from the whole source dictionary for testing purposes. The testing lexicon does not contain verbs *být* (be) and modal verbs as they need special treatment. The results of the automatic procedure on this portion of the lexicon are shown below.

D.3.1. Verbs processed fully automatically

```
brát    R--s[i1]1(hPTc1)2[hPTc4|hPc4r{na}]%$
čekat   RSEs[i1]1[hPTc1]@
činit   RSIs[i1]1[hPTc1]@
činit   R--s[i1]1(hPTc1)2[hTc4]3(hPc3)%$
dát     RSEs[i1]1[hPTc1]A[sI|hPTc2r{do}|hPTc3r{ke}|hTc4r{na}]@
dát     R--s[i1]1(hPTc1)2[hPTc4]3[hPc3]%$
dát     R--s[i1]1(hPTc1)2[sP|sU|hPc4r{na}]%$
dělat   RSEs[i1]1[hPTc1]2[hTc4]4[hPc2r{z}]@
dělat   R--s[i1]1(hPTc1)A[hPTc4r{na}|hTc2r{do}]
          EJ[hTc6r{na}|hTc6r{v}]V[hPTc7r{s}|hTc6r{v}]$
dělat   R--s[i1]1(hPTc1)2[hPTc4]3(hPc3|hRc2r{ze})%$
dít     RSEs[i1]1[hPTc1]2[hPc3]A[hA]@
dít     R--s[i1]1(hPTc1)2[hTc4]3[hPc3]%$
dodat   R--s[i1]1(hPTc1)2[hTc2]3[hPTc3]%$
dodat   R--s[i1]1(hPTc1)2[hTc4]A[hTc2r{do}|hTc3r{k}]%$
dodat   R--s[i1]1(hPTc1)2[hTc4]3(hPc3|hTc3r{k})A(hTc2r{do}|hTc3r{k})
          %$
dojít   R--s[i1]1(hPTc1)2[hPTc2|hPTc4r{pro}|hTc3r{k}]
          A[hPTc3r{ke}|hPTc4r{na}|hTc3r{k}]%$
dojít   R--s[i1]1(hPTc1)2[hTc4]3[hPc3]%$
dokázat R--s[i1]1(hPTc1)2[sD|sI]%$
dosáhnout       R--s[i1]1(hPTc1)A[hTc2r{do}|hTc3r{k}|hTc4r{na}]$
```

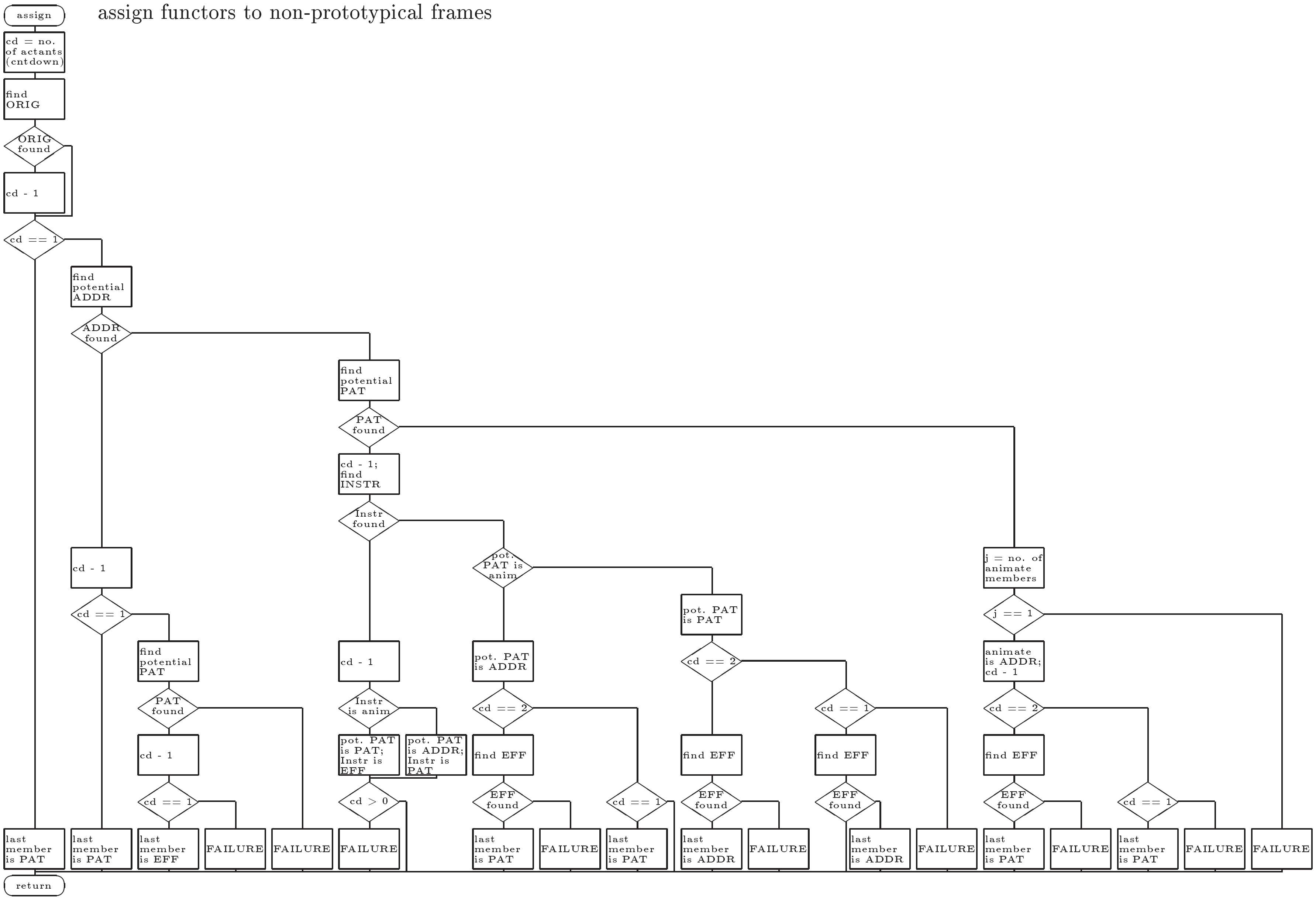


Figure D.1.: The algorithm for assigning functors to non-prototypical frames

```
dosáhnout       R--s[i1]1(hPTc1)2[hTc2]4(hPc6r{na})%$
dostat  RSEs[i1]1[hPTc1]A[hPTc3r{ke}|hPTc4r{na}|hPTc4r{přes}]@
dostat  R--s[i1]1(hPTc1)2[hPTc4|hTc4]A[hTc2r{do}|hTc4r{na}]
           O[hPTc2r{z}|hTc2r{z}]%$
existovat       R--s[i1]1(hPTc1)E[hTc2r{u}|hTc6r{na}|hTc6r{v}]$
hledat  R--s[i1]1(hPTc1)2[hPTc4]E[hPTc2r{u}|hPTc6r{v}|hTc6r{na}]
           %$
hrát    R--s[i1]1(hPTc1)2[hTc4]3(hPc3|hPc4r{pro})%$
chybět  R--s[i1]1(hPTc1)2[hPTc1]EJ[hTc6r{na}|hTc6r{v}]%$
chybět  R--s[i1]1(hPTc1)2[hPTc1]3(hPc3)EJ(hTc6r{na}|hTc6r{v})%$
chybět  R--s[i1]1(hPTc1)2[hTc4]E[hTc2r{u}]%$
informovat      R--s[i1]1(hPTc1)2(hPTc6r{o})3[hPc4]%$
jednat  RSEs[i1]1[hPTc1]2[hPTc4r{o}]@
jednat  R--s[i1]1(hPTc1)2[hPTc4]%$
mít     R--s[i1]1(hPTc1)2[hPTc4]A(hTc3r{k}|hTc4r{za})%$
nabízet R--s[i1]1(hPTc1)2[hPTc4]3(hPc3)%$
najít   RSEs[i1]1[hPTc1]E[hPc2r{u}|hTc6r{na}|hTc6r{v}]@
najít   R--s[i1]1(hPTc1)2[hTc4]4[hPc6r{na}]%$
napsat  RSEs[i1]1[hPTc1]A[hTc2r{do}|hTc4r{na}]@
napsat  RSIs[i1]1[hPTc1]2[hTc4]A(hTc2r{do}|hTc4r{na})@
nechat  RSIs[i1]1[hPTc1]2[hTc4|sI]@
nechat  R--s[i1]1(hPTc1)2[hPTc2]%$
nechat  R--s[i1]1(hPTc1)2(sI|hPc3)3[hPTc4]
           E(hPc2r{u}|hTc6r{na}|hTc6r{v})%$
nechat  R--s[i1]1(hPTc1)2[hTc4]4(hPc6r{na})%$
objevit RSEs[i1]1[hPTc1]E[hPc2r{u}|hTc6r{na}|hTc6r{při}|hTc6r{v}]
           @
objevit R--s[i1]1(hPTc1)2[hPTc4]EJ(hTc6r{na}|hTc6r{v})%$
očekávat        R--s[i1]1(hPTc1)2[hPTc4]EJ(hTc6r{na}|hTc6r{v})%$
odmítnout       R--s[i1]1(hPTc1)2[hPTc4]%$
odmítnout       R--s[i1]1(hPTc1)2[hTc4]3[hPc3]%$
otevřít RSEs[i1]1[hPTc1]2[hPTc3]@
oznámit R--s[i1]1(hPTc1)2[sD|sR|sZ]%$
patřit  RSEs[i1]1[hPTc1]2[sI|sU]@
patřit  R--s[i1]1(hPTc1)2[hPc3]
           A[hPTc3r{ke}|hPTc4r{mezi}|hPTc4r{pod}|hTc2r{do}|hTc4r{na}]%$
počítat R--s[i1]1(hPTc1)2[sD|sR|sZ]%$
podařit RSEs[i1]1[hPTc1]2[hPc3]@
podepsat        RSEs[i1]1[hPTc1]EJ[hTc6r{na}|hTc6r{v}]@
podepsat        R--s[i1]1(hPTc1)2[hPTc4]%$
podepsat        R--s[i1]1(hPTc1)2[hTc4]3[hPc3]%$
pomoci  R--s[i1]1(hPTc1)2[hPc3]A(hTc2r{do}|hTc3r{k})
           E(hTc6r{při}|hTc6r{v})O(hPTc2r{od}|hTc2r{z})%$
potřebovat      R--s[i1]1(hPTc1)2[hPTc4]AP(hTc3r{k}|hTc4r{na})%$
potřebovat      R--s[i1]1(hPTc1)2[hTc2|sI|sU]%$
potvrdit        RSEs[i1]1[hPTc1]2[sD|sR|sZ]@
```

```
potvrdit        R--s[i1]1(hPTc1)2[hPTc4|sD|sR|sZ]%$
považovat       RSIs[i1]1[hPTc1]2[hPTc2]@
považovat       R--s[i1]1(hPTc1)$
pracovat        R--s[i1]1(hPTc1)2[hPTc4r{pro}|hTc3r{k}]
          A[hPTc4r{na}|hTc3r{k}|hTc4r{přes}]
          E[hPTc2r{u}|hPTc6r{na}|hTc6r{v}]I[hTc7|hTc4r{přes}]
          J[hTc4r{jako}|hTc4r{přes}|hTc6r{v}]%$
prodat  R--s[i1]1(hPTc1)2(hPTc4)3[hPc3]
          A(hTc4r{na}|hTc4r{za}|hTc6r{po})%$
prohlásit       RSEs[i1]1[hPTc1]S[hPTc7|hPTc4r{za}]@
prohlásit       R--s[i1]1(hPTc1)2(hPTc6r{o})5[hTc4]%$
předpokládat    R--s[i1]1(hPTc1)2[sD]%$
představovat    RSIs[i1]1[hPTc1]2[sD|sR|sZ]@
přijít  RSIs[i1]1[hPTc1]A[sI|hPTc4r{na}]@
přijmout        R--s[i1]1(hPTc1)2[hPc4]J(hPTc4r{jako}|hTc4r{za})%$
přinést R--s[i1]1(hPTc1)2[hPTc4]EJ(hTc6r{na}|hTc6r{v})%$
připravit       RSIs[i1]1[hPTc1]2[hPTc4]A(hPTc4r{na}|hTc2r{do}|hTc3r{k})
          @
připravit       R--s[i1]1(hPTc1)A[hTc2r{do}|hTc3r{k}]$
připravit       R--s[i1]1(hPTc1)2[hPc4]
          A[hTc2r{do}|hTc3r{k}|hTc4r{na}|hTc4r{o}]%$
připravit       R--s[i1]1(hPTc1)2[hTc4]3(hPc3)
          A(hTc2r{do}|hTc3r{k}|hTc4r{na})%$
působit R--s[i1]1(hPTc1)2[hTc4]3[hPc3]%$
rozhodnout      R--s[i1]1(hPTc1)2[hTc4]J[hTc6r{o}|hTc6r{v}]%$
rozhodnout      R--s[i1]1(hPTc1)2[sD|sR|sZ]%$
rozpočíst       RSEs[i1]1[hPTc1]@
říci    RSIs[i1]1[hPTc1]2(hPTc6r{o})5[hTc4]@
říci    RSIs[i1]1[hPTc1]2[sD|sR|sZ]@
říci    R--s[i1]1(hPTc1)2[hTc4]3(hPTc3|hPTc6r{o}|hTc3r{k})
          A(hPTc2r{do}|hPTc4r{na}|hTc3r{k})P(hTc3|hTc3r{k})%$
říci    R--s[i1]1(hPTc1)2[hPTc6r{o}|hTc3r{k}]5[hTc4]
          A[hPTc2r{do}|hPTc4r{na}|hTc3r{k}]P[hTc3|hTc3r{k}]%$
říci    R--s[i1]1(hPTc1)2[sD|sR|sZ]3(hPc3)%$
říci    R--s[i1]1(hPTc1)2[sP|sU]%$
říkat   RSIs[i1]1[hPTc1]2(hPTc6r{o})5[hTc4]@
říkat   RSIs[i1]1[hPTc1]2[sD|sR|sZ]@
říkat   R--s[i1]1(hPTc1)2[hTc4]3(hPTc3|hPTc6r{o}|hTc3r{k})
          A(hPTc2r{do}|hPTc4r{na}|hTc3r{k})P(hTc3|hTc3r{k})%$
říkat   R--s[i1]1(hPTc1)2[hPTc6r{o}|hTc3r{k}]5[hTc4]
          A[hPTc2r{do}|hPTc4r{na}|hTc3r{k}]P[hTc3|hTc3r{k}]%$
říkat   R--s[i1]1(hPTc1)2[sD|sR|sZ]3(hPc3)%$
říkat   R--s[i1]1(hPTc1)2[sP|sU]%$
sdělit  RSIs[i1]1[hPTc1]2[sD|sR|sZ]@
sdělit  R--s[i1]1(hPTc1)2[sD|sR|sZ]%$
skončit RSEs[i1]1[hPTc1]@
```

```
skončit R--s[i1]1(hPTc1)E[hTc2r{u}|hTc6r{na}|hTc6r{v}]
           V[hPTc7r{s}|hTc6r{v}]$
slyšet  R--s[i1]1(hPTc1)2(hPTc6r{o})5[hTc4]%$
snažit  RSEs[i1]1[hPTc1]2[sI|sU]@
stanovit        R--s[i1]1(hPTc1)2[hTc4]3(hPc3)EJ(hTc6r{na}|hTc6r{v})%$
stát    RSIs[i1]1[hPTc1]E[hTc6r{na}|hTc7r{za}]@
stát    R--s[i1]1(hPTc1)
           2[sD|hM|hPc3r{proti}|hPc6r{na}|hPTc4r{o}|hPTc6r{při}]
           A[hPc7r{za}|hPc3r{proti}|hTc4r{za}|hTc7r{za}]
           E[hPTc7r{nad}|hTc2r{kolem}|hTc2r{u}|hTc6r{v}|hTc7r{pod}|hTc7r{za}]
           P[hPTc7r{před}|hTc3r{proti}]%$
trvat   R--s[i1]1(hPTc1)EJ[hTc6r{na}|hTc6r{v}]$
tvrdit  R--s[i1]1(hPTc1)2[hTc4]3(hPc3|hPTc6r{o})%$
tvrdit  R--s[i1]1(hPTc1)2[sD]%$
udělat  RSEs[i1]1[hPTc1]EJ[hTc6r{na}|hTc6r{v}]@
udělat  RSIs[i1]1[hPTc1]2[hTc4]@
ukázat  R--s[i1]1(hPTc1)2[hPTc4]3[hPc3]%$
ukázat  R--s[i1]1(hPTc1)2[sD|sR|sZ]A[hPTc4r{na}|hTc2r{do}]%$
uvádět  R--s[i1]1(hPTc1)2[hTc4]3[hPc3]%$
uvést   R--s[i1]1(hPTc1)2[hPc4]
           A(hPc3r{ke}|hTc2r{do}|hTc4r{na}|hTc4r{v})%$
uvést   R--s[i1]1(hPTc1)2[hPTc4]EJ(hTc6r{na}|hTc6r{v})%$
uvést   R--s[i1]1(hPTc1)2[hPTc4]3(hPc3)EJ(hTc6r{na}|hTc6r{v})%$
vědět   R--s[i1]1(hPTc1)2[sD|sR|sZ]%$
věnovat RSEs[i1]1[hPTc1]2[hPTc3]@
věřit   R--s[i1]1(hPTc1)2[hTc4]3[hPc3]%$
vést    R--s[i1]1(hPTc1)2[hPTc4]%$
vrátit  R--s[i1]1(hPTc1)2[hTc4]3[hPc3]%$
vydat   R--s[i1]1(hPTc1)2[hPTc4]A(hTc4r{na}|hTc4r{v})%$
vydat   R--s[i1]1(hPTc1)2[hPTc4]3(hPTc3)A(hTc4r{na}|hTc4r{v})%$
vyhrát  R--s[i1]1(hPTc1)C[hTc7|hPTc7r{nad}]$
vyhrát  R--s[i1]1(hPTc1)2[hTc4]4[hPc6r{na}]%$
vycházet        R--s[i1]1(hPTc1)A[hTc2r{do}|hTc4r{na}]
           O[hPTc2r{z}|hTc2r{od}]$
vzít    RSIs[i1]1[hPTc1]2[hPc4]A[hTc4r{na}|hTc4r{za}]@
vzít    R--s[i1]1(hPTc1)A[hTc2r{do}|hTc3r{k}|hTc4r{na}]
           EJU[hTc2r{kolem}]$
vzít    R--s[i1]1(hPTc1)2[hPTc4]3[hPc3]%$
začínat R--s[i1]1(hPTc1)2[sI|hPTc6r{o}]O[hTc2r{od}|hTc2r{z}]%$
začít   R--s[i1]1(hPTc1)2[sI|hPTc6r{o}]O[hTc2r{od}|hTc2r{z}]%$
zdát    RSEs[i1]1[hPTc1]2[sD|sZ|hQc1|hPTc6r{o}]3(hPc3)@
získat  RSIs[i1]1[hPTc1]2[hTc4]@
zjistit RSIs[i1]1[hPTc1]2[hTc4]E(hPc2r{u}|hTc6r{na}|hTc6r{v})@
zjistit R--s[i1]1(hPTc1)2[hPTc4]%$
zjistit R--s[i1]1(hPTc1)2[hTc4]E(hPc2r{u}|hTc6r{na}|hTc6r{v})%$
změnit  RSEs[i1]1[hPTc1]A[hTc4r{na}|hTc4r{v}]@
```

```
znamenat        R--s[i1]1(hPTc1)2[hTc4|sD]%$
znát     R--s[i1]1(hPTc1)2[v{eN}hTc2|hPTc4]%$
zůstat   R--s[i1]1(hPTc1)E[hPTc2r{u}|hTc6r{na}|hTc6r{v}]$
zvýšit   R--s[i1]1(hPTc1)2[hTc4]3(hPc3)EJ(hTc6r{na}|hTc6r{v})
            F(hMr{o}|hTc4r{o})%$
žít      RSIs[i1]1[hPTc1]EJ[hTc6r{na}|hTc6r{v}]V[hPc7r{s}|hTc6r{v}]@
žít      R--s[i1]1(hPTc1)2[hTc4]%$
```

D.3.2. Verbs with ambiguous frames

```
brát     RSEs[i1]1[hPTc1]2[hPTc4r{o}]2O[hTc2r{z}]2EJV[hTc6r{v}]@
brát     RSIs[i1]1[hPTc1]2[hPc4]2ACJ(hTc4r{za})@
brát     RSIs[i1]1[hPTc1]2[hTc4]2A[hTc2r{do}]2CI(hTc7)2O(hTc2r{z})@
brát     R--s[i1]1(hPTc1)2[hPc4|hTc4]2ACJ[hTc4r{za}]%$
brát     R--s[i1]1(hPTc1)2[hTc4]2[hPc3]2O[hTc2r{z}]2ACJ[hTc4r{za}]%$
cítit    RSEs[i1]1[hPTc1]2CI[hTc7]2AP[hTc4r{na}]@
cítit    R--s[i1]1(hPTc1)2[hPTc4]2EJV(hTc6r{v})%$
cítit    R--s[i1]1(hPTc1)2[hTc4]2CI(hTc7)%$
cítit    R--s[i1]1(hPTc1)2LV[hPc7r{s}]$
čekat    R--s[i1]1(hPTc1)2A[hPTc4r{na}]$
čekat    R--s[i1]1(hPTc1)2[hPc3]2IV[hTc7r{s}]%$
čekat    R--s[i1]1(hPTc1)2[hPTc4|hTc4]2LV[hPc7r{s}]%$
čekat    R--s[i1]1(hPTc1)2[hPTc4]2LV(hPc7r{s})%$
činit    RSEs[i1]1[hPTc1]2EJV[hTc6r{v}]@
dát      RSIs[i1]1[hPTc1]2[hTc4]2IV(hTc7r{s})@
dát      R--s[i1]1(hPTc1)2A[sI]2[hPc3]%$
dát      R--s[i1]1(hPTc1)2[hPTc4]2AP[hTc4r{na}]2A[hTc2r{do}]
            A[hTc4r{před}]%$
dělat    RSEs[i1]1[hPTc1]2IS[hPTc7]2V[hPTc7r{s}]@
dělat    R--s[i1]1(hPTc1)2[hTc4]2[hM|hPc6r{po}|hPc4r{pro}]
            2IV[hTc7r{s}]E[hPc2r{u}]%$
dojít    R--s[i1]1(hPTc1)2LV[hPc7r{s}]2A(hTc2r{do})$
dokázat R--s[i1]1(hPTc1)2[hTc4]2(hPc3)2EJV(hTc6r{v})%$
dosáhnout       R--s[i1]1(hPTc1)2[hTc4]2CI(hTc7)%$
dostat   RSEs[i1]1[hPTc1]2CI[hTc7]2O(hTc2r{z})A(hTc2r{do}|hTc4r{na})
            @
dostat   RSEs[i1]1[hPTc1]2[hTc2]2A[hTc2r{do}]2(hPc3)@
dostat   R--s[i1]1(hPTc1)2CI[hTc7]%$
dostat   R--s[i1]1(hPTc1)2CI[hTc7]A(hTc2r{do}|hTc4r{na}|hTc6r{po})
            %$
dostat   R--s[i1]1(hPTc1)2[hPTc4]2O(hTc2r{z})A(hTc2r{do}|hTc4r{na})
            %$
dostat   R--s[i1]1(hPTc1)2[hTc4]2ACJ[hTc4r{za}]
            O[hPc2r{od}|hPTc2r{z}]%$
dostat   R--s[i1]1(hPTc1)2[hTc4]2O[hPTc2r{z}]%$
hovořit R--s[i1]1(hPTc1)2CI[hTc7]2A(hPc3r{ke})%$
```

```
hovořit R--s[i1]1(hPTc1)2LV[hPc7r{s}]2JY(hTc6r{o})$
hrát    RSEs[i1]1[hPTc1]2V[hPTc7r{s}]@
hrát    RSIs[i1]1[hPTc1]2A[hPTc4r{na}]2V(hPTc7r{s})@
hrát    R--s[i1]1(hPTc1)2[hPc3]2A(hTc2r{do})%$
hrát    R--s[i1]1(hPTc1)2[hPTc4]2EJV(hTc6r{v})%$
hrát    R--s[i1]1(hPTc1)2V[hPTc7r{s}]A[hTc4r{na}|hTc4r{o}]$
chodit  R--s[i1]1(hPTc1)2AF[hTc4r{o}]2CP[hTc4r{pro}]
           CP[hTc3r{kvůli}]$
chodit  R--s[i1]1(hPTc1)2A[hPc3r{ke}]A(hTc4r{na}|hTc4r{o})
           CP(hTc3r{kvůli})$
chodit  R--s[i1]1(hPTc1)2B[hPc4r{na}]2IV(hTc7r{s})
           A[hPc7r{za}|hTc4r{na}]$
chodit  R--s[i1]1(hPTc1)2B[hPc4r{pro}]2A(hTc2r{do})$
chodit  R--s[i1]1(hPTc1)2CP[hTc4r{pro}]
           A[sI|hPc3r{ke}|hPTc7r{za}|hTc4r{o}]V[hPTc7r{s}|hTc6r{v}]%$
informovat     RSEs[i1]1[hPTc1]2[hPTc6r{o}]2AP[hTc4r{na}]E[hPc2r{u}]
           @
jednat  R--s[i1]1(hPTc1)2LV[hPc7r{s}]2(hPTc6r{o})2AF(hTc4r{o})$
jet     R--s[i1]1(hPTc1)2CI[hTc7]AJTUY(hTc6r{po})%$
jet     R--s[i1]1(hPTc1)2[hTc4|hPc3|hPc4r{pro}]2E[hPTc6r{na}]
           2IV[hTc7r{s}]
           A[hPc7r{za}|hPc3r{ke}|hPTc6r{po}|hTc2r{do}|hTc4r{na}]%$
jít     R--s[i1]1(hPTc1)2A[hPc3r{ke}]CP(hTc3r{kvůli}|hTc4r{pro})$
jít     R--s[i1]1(hPTc1)2A[hPTc3r{ke}]2AF(hTc4r{o})2IV(hTc7r{s})$
jít     R--s[i1]1(hPTc1)2A[hTc2r{do}]2CP(hTc4r{pro})$
jít     R--s[i1]1(hPTc1)
           2[hTc4|sI|hPTc3r{proti}|hPTc4r{o}|hPTc4r{pro}|hTc3r{k}]
           2S[hPTc4r{za}]C[hTc7|hTc2r{od}]%$
jít     R--s[i1]1(hPTc1)2IV[hTc7r{s}]A[hPc7r{za}|hPTc4r{na}]$
mít     RSEs[i1]1[hPTc1]2A[hPTc3r{ke}]@
mít     R--s[i1]1(hPTc1)2A[sI]A(hA)%$
mít     R--s[i1]1(hPTc1)2[hTc4]2O(hTc2r{z})A(sI|hTc4r{za})
           V(hPc7r{s}|hTc6r{v})%$
mluvit  R--s[i1]1(hPTc1)2[hPc3]2A[hTc2r{do}]%$
mluvit  R--s[i1]1(hPTc1)
           2[hPTc4|hPc2r{do}|hPc4r{na}|hPTc3r{proti}|hPTc4r{pro}]
           2A[hPTc3r{ke}]2S[hPTc4r{za}]%$
mluvit  R--s[i1]1(hPTc1)2LV[hPc7r{s}]2(hPTc6r{o})$
myslit  R--s[i1]1(hPTc1)2[hPTc4|sD|hPTc6r{o}]2A[hPTc4r{na}]
           2EJV[hTc6r{v}]%$
nabízet RSEs[i1]1[hPTc1]2[hPc3]2AP(hTc3r{k})@
najít   R--s[i1]1(hPTc1)2[hPTc4]2O(hTc2r{z})EJ(hTc6r{na}|hTc6r{v})
           %$
napsat  RSIs[i1]1[hPTc1]2[hPc3]2AF(hTc4r{o})@
napsat  R--s[i1]1(hPTc1)2[hTc4]2(hPc3)2CI[hTc7]
           A(hTc2r{do}|hTc4r{na})%$
```

```
nechat  RSEs[i1]1[hPTc1]2A[sI]@
očekávat        R--s[i1]1(hPTc1)2[hTc4]2O[hPTc2r{od}]%$
odpovídat       RSEs[i1]1[hPTc1]2[hPc3]2O(hTc2r{z})@
odpovídat       R--s[i1]1(hPTc1)2[hPc3]2ACJ(hTc4r{za})%$
odpovídat       R--s[i1]1(hPTc1)2[hPc3]2CI[hTc7]2AP[hTc4r{na}]%$
odpovídat       R--s[i1]1(hPTc1)2P[hTc3]%$
otevřít RSEs[i1]1[hPTc1]2[hPTc3]2E[hPc7r{před}]@
otevřít R--s[i1]1(hPTc1)2[hTc4]2(hPc3)2CI(hTc7)2A(hTc2r{do})%$
otevřít R--s[i1]1(hPTc1)2[hTc4]2(hPTc3)2CI(hTc7)2A(hTc2r{do})%$
oznámit R--s[i1]1(hPTc1)2[hTc4]2[hPc3]2CI(hTc7)2B(hPc4r{na})%$
pět     R--s[i1]1(hPTc1)2[hTc4]2JY(hTc6r{o})%$
platit  R--s[i1]1(hPTc1)2[hPc3]2[hTc4]2CI[hTc7]2ACJ[hTc4r{za}]%$
platit  R--s[i1]1(hPTc1)2[hPTc3]2A[hPTc4r{na}]2S[hPTc4r{za}]
           2CDOZ[hTc2r{od}]E[hPc2r{u}]%$
platit  R--s[i1]1(hPTc1)2[hPTc4]2ACJ(hTc4r{za})%$
platit  R--s[i1]1(hPTc1)2[hTc4]2B[hPTc4r{pro}]%$
počíst  RSIs[i1]1[hPTc1]2[hTc4|hTc6r{o}]2EJV[hTc6r{v}]@
počítat R--s[i1]1(hPTc1)2[hPTc4]2ACJ(hTc4r{za})%$
počítat R--s[i1]1(hPTc1)2V[hPTc7r{s}]P(hTc4r{na}|hTc4r{pro})$
podepsat        RSEs[i1]1[hPTc1]2[hPc3]2CI(hTc7)A(hTc2r{do}|hTc4r{na})@
pokračovat      R--s[i1]1(hPTc1)2CI[hTc7]
           A[hPTc3r{ke}|hTc2r{do}|hTc4r{na}]EJ[hTc6r{na}|hTc6r{v}]%$
pomoci  RSIs[i1]1[hPTc1]2CI[hTc7]2O[hPTc2r{od}]2AP[hTc3r{k}]
           E[hTc6r{při}|hTc6r{v}]V[hPTc7r{s}|hTc6r{v}]@
pomoci  R--s[i1]1(hPTc1)2[hPTc3]2CI(hTc7)%$
potvrdit        R--s[i1]1(hPTc1)2[hTc4]2[hPc3]2CI[hTc7]%$
pracovat        R--s[i1]1(hPTc1)2V[hPTc7r{s}]2EJ(hTc6r{na})$
prodat  RSEs[i1]1[hPTc1]2[hPc3]2ACJ(hTc4r{za})@
prohlásit       R--s[i1]1(hPTc1)2[hPc4]2CI(hTc7)%$
prohlásit       R--s[i1]1(hPTc1)2[hPTc4]2S(hPTc4r{za})%$
předpokládat    R--s[i1]1(hPTc1)2[hTc4]2JY(hTc6r{o})%$
představovat    RSEs[i1]1[hPTc1]2[hPc3]2CI(hTc7)@
představovat    RSIs[i1]1[hPTc1]2[hPTc4]2JR(hTc4r{jako})@
představovat    R--s[i1]1(hPTc1)2[hPTc4]2(hPc3)2CI(hTc7)
           EJ(hTc6r{na}|hTc6r{v})%$
představovat    R--s[i1]1(hPTc1)2[hTc4]2P(hTc3)%$
přijít  R--s[i1]1(hPTc1)2A[hPc3r{ke}]2B[hPc4r{na}]2IV(hTc7r{s})$
přijít  R--s[i1]1(hPTc1)2[sI|hPc3|hPTc4r{o}|hPTc4r{pro}]
           2V[hPTc7r{s}]%$
přijmout        R--s[i1]1(hPTc1)2[hPc4|hTc4]2J[hPTc4r{jako}]
           2AP[hTc4r{na}]%$
přijmout        R--s[i1]1(hPTc1)2[hPTc4]2J(hPc7)A(hTc2r{do}|hTc4r{na})
           E(hPc2r{u})%$
přijmout        R--s[i1]1(hPTc1)2[hTc4]2IV[hTc7r{s}]%$
přijmout        R--s[i1]1(hPTc1)2[hTc4]2O(hPc2r{od})2AP(hTc4r{na})%$
přijmout        R--s[i1]1(hPTc1)2J[hPTc4r{jako}]$
```

```
přinést R--s[i1]1(hPTc1)2[hPTc4]2[hPTc3]2CI[hTc7]2O[hTc2r{z}]
           A[hTc2r{do}|hTc4r{na}]%$
připravit       RSEs[i1]1[hPTc1]2CI[hTc7]2AF(hTc4r{o})@
připravit       RSEs[i1]1[hPTc1]2CI[hTc7]2EJV[hTc6r{v}]
           A[hPTc4r{na}|hTc2r{do}|hTc3r{k}]@
připravit       R--s[i1]1(hPTc1)2[hPc4]2AP(hTc3r{k})%$
psát    RSEs[i1]1[hPTc1]2JR[hTc4r{jako}]@
psát    RSIs[i1]1[hPTc1]2LV[hPc7r{s}]2JY(hTc6r{o})
           A(hTc2r{do}|hTc4r{na})@
psát    R--s[i1]1(hPTc1)2B[hPTc4r{pro}]EJ[hTc6r{na}|hTc6r{v}]$
psát    R--s[i1]1(hPTc1)2[hTc4]2A[hPTc4r{na}]%$
psát    R--s[i1]1(hPTc1)2[hTc4]2CI(hTc7)2(hPc3)2JY(hTc6r{o})%$
působit R--s[i1]1(hPTc1)2CI[hTc7]2A(hPTc4r{na})%$
působit R--s[i1]1(hPTc1)2P[sU]2JR[hTc4r{jako}]%$
rozhodnout      RSEs[i1]1[hPTc1]
           2[sD|sR|sZ|hPTc4r{pro}|hTc3r{k}|hTc6r{o}]2EU[hPTc7r{mezi}]@
rozpočíst       R--s[i1]1(hPTc1)2CI[hTc7]%$
rozpočíst       R--s[i1]1(hPTc1)2[hTc4]2AP(hTc4r{na})%$
růst    R--s[i1]1(hPTc1)2[hPc3]2CI[hTc7]2A[hTc2r{do}]2O[hTc2r{z}]
           EJ[hTc6r{na}|hTc6r{v}]%$
říci    RSIs[i1]1[hPTc1]2[hPc3]2AF(hTc4r{o})@
říci    R--s[i1]1(hPTc1)2[sD|sR|sZ]2AF[hTc4r{o}]2[hPc3]%$
říkat   RSIs[i1]1[hPTc1]2[hPc3]2AF(hTc4r{o})@
říkat   R--s[i1]1(hPTc1)2[sD|sR|sZ]2AF[hTc4r{o}]2[hPc3]%$
sdělit  RSIs[i1]1[hPTc1]2[hTc4]2LV[hPc7r{s}]@
sdělit  R--s[i1]1(hPTc1)2[hTc4]2[hPTc3]2JY(hTc6r{o})%$
skončit R--s[i1]1(hPTc1)2[hTc4]2CI(hTc7)%$
slyšet  R--s[i1]1(hPTc1)2[hPTc4|sD|sR|sZ]2AP[hTc4r{na}]%$
smět    R--s[i1]1(hPTc1)2A[sI]%$
stačit  R--s[i1]1(hPTc1)2[hPc3]2AP(hTc3r{k})%$
stačit  R--s[i1]1(hPTc1)2[hPTc3]2CI(hTc7)2EJV(hTc6r{v})%$
stačit  R--s[i1]1(hPTc1)2IV[hTc7r{s}]A[sI|hPTc4r{na}]
           BP[hTc3r{proti}]%$
stanovit        RSIs[i1]1[hPTc1]2[hTc4]2JR(hTc4r{jako})@
stát    RSEs[i1]1[hPTc1]2[sD|hQc7|hPTc3]2IS[hPTc7]@
stát    R--s[i1]1(hPTc1)2[hPc3]2EJ(hTc6r{na})2CI(hTc7)
           AJTUY(hTc6r{po})%$
tvořit  RSEs[i1]1[hPTc1]2CI[hTc7]2O[hTc2r{z}]@
tvořit  R--s[i1]1(hPTc1)2[hTc4]2CI(hTc7)2O(hTc2r{z})%$
udělat  R--s[i1]1(hPTc1)2[hPc4]2[hQc7]2IS[hPTc7]%$
udělat  R--s[i1]1(hPTc1)2[hTc4]2CI(hTc7)2O(hPTc2r{z})2V(hPTc7r{s})
           %$
udělat  R--s[i1]1(hPTc1)2[hTc4]2[hPc3]2O[hTc2r{z}]AIJU[hTc4r{přes}]
           %$
ukázat  RSEs[i1]1[hPTc1]2[sD|sR|sZ]2P[hPTc7r{před}]
           J[hTc4r{jako}|hTc6r{na}|hTc6r{v}]@
```

```
uvádět  RSEs[i1]1[hPTc1]2CI[hTc7]EJ(hTc6r{na}|hTc6r{v})@
uvádět  R--s[i1]1(hPTc1)2[hPc4]2A(hPc3r{ke})%$
uvádět  R--s[i1]1(hPTc1)2[hPTc4]2CI(hTc7)%$
uvádět  R--s[i1]1(hPTc1)2[hPTc4]2CI(hTc7)
           A(hTc2r{do}|hTc4r{na}|hTc4r{v})%$
uvést   RSEs[i1]1[hPTc1]2CI[hTc7]2EJV[hTc6r{v}]@
uvést   R--s[i1]1(hPTc1)2[hTc4]2CI(hTc7)2A(hTc2r{do})%$
vědět   R--s[i1]1(hPTc1)2[hTc4]2O[hPc2r{od}]2[hPc4r{na}|hPTc6r{o}]
           %$
věnovat R--s[i1]1(hPTc1)2[hTc4]2[hPTc3]2A[hPTc4r{na}]%$
věřit   R--s[i1]1(hPTc1)2[sD|sR|sZ|hPTc3|hPTc4r{v}]2A[hPTc4r{na}]
           %$
vést    RSEs[i1]1[hPTc1]2[hPTc3]2LV[hPc7r{s}]@
vést    RSIs[i1]1[hPTc1]2EJV[hTc6r{v}]@
vést    R--s[i1]1(hPTc1)2CE[hPc7r{nad}]2EJV(hTc6r{v})$
vést    R--s[i1]1(hPTc1)2[hPc4]2AP[hTc3r{k}]%$
vést    R--s[i1]1(hPTc1)2[hTc4]2LV(hPc7r{s})
           A(hTc2r{do}|hTc4r{na}|hTc4r{přes})I(hTc7|hTc4r{přes})%$
vidět   RSEs[i1]1[hPTc1]2E[hPTc6r{v}]@
vidět   RSEs[i1]1[hPTc1]2LV(hPc7r{s})E[hPc2r{u}|hTc6r{na}|hTc6r{v}]
           @
vidět   R--s[i1]1(hPTc1)2[hPc3]2A[hTc2r{do}]%$
vidět   R--s[i1]1(hPTc1)2[hPc4|hPTc4]2E[hPTc6r{v}]%$
vidět   R--s[i1]1(hPTc1)2[hPTc4]2CI(hTc7)2P(hPTc7r{před})%$
vidět   R--s[i1]1(hPTc1)2[hPTc4]2E[hPTc6r{v}]%$
vidět   R--s[i1]1(hPTc1)2[hTc4]2E[hPTc6r{na}]%$
vidět   R--s[i1]1(hPTc1)2[sD|sR|sZ]2A[hPTc4r{na}]%$
vrátit  RSEs[i1]1[hPTc1]2[hPc3|hTc3r{k}]2V[hPTc7r{s}]2O[hTc2r{z}]
           A[hTc2r{do}|hTc3r{k}|hTc4r{na}]@
vrátit  R--s[i1]1(hPTc1)2[hPTc4]2O(hTc2r{z})
           A(hTc2r{do}|hTc3r{k}|hTc4r{na})P(hTc3|hTc3r{k}|hTc4r{na})%$
vydat   RSEs[i1]1[hPTc1]2[hPc3]2CI[hTc7]2S[hPTc4r{za}]2O[hTc2r{z}]
           A[hPc7r{za}|hPc3r{ke}|hTc2r{do}|hTc4r{na}|hTc4r{v}]@
vydat   R--s[i1]1(hPTc1)2[hPc4|hTc4]2S(hPTc4r{za})%$
vydat   R--s[i1]1(hPTc1)2[hTc4]2CI(hTc7)2A(hPTc4r{na})2S(hPTc4r{za})
           %$
vydat   R--s[i1]1(hPTc1)2[hTc4]2(hPc3)2CI(hTc7)2A(hPTc4r{na})
           2S(hPTc4r{za})%$
vyhrát  RSEs[i1]1[hPTc1]2EJV[hTc6r{v}]@
vyhrát  RSIs[i1]1[hPTc1]2[hTc4]2EJ[hTc6r{na}]2IV[hTc7r{s}]@
vyhrát  R--s[i1]1(hPTc1)2V[hPTc7r{s}]2EJV(hTc6r{v})$
vycházet        R--s[i1]1(hPTc1)2[hTc4]2A[hPTc4r{na}]%$
vycházet        R--s[i1]1(hPTc1)2[hTc4]2[hPc3]2EJV(hTc6r{v})%$
vycházet        R--s[i1]1(hPTc1)2IV(hTc7r{s})A[hTc2r{do}|hTc4r{na}]$
vycházet        R--s[i1]1(hPTc1)2LV[hPc7r{s}]2EJV(hTc6r{v})$
vycházet        R--s[i1]1(hPTc1)2O[hTc2r{z}]E(hTc6r{při}|hTc6r{v})$
```

```
vyjít   RSIs[i1]1[hPTc1]2[hTc4]2O[hTc2r{z}]
          A[hTc2r{do}|hTc4r{na}|hTc6r{po}]@
vyjít   R--s[i1]1(hPTc1)2[hM]2A(hPTc4r{na})$
vyjít   R--s[i1]1(hPTc1)2[hTc4|sD|hM]2AP[hTc4r{na}]2[hPc3]%$
vyjít   R--s[i1]1(hPTc1)2[hTc4]2EJV[hTc6r{v}]%$
vyjít   R--s[i1]1(hPTc1)2V[hPTc7r{s}]A[hTc2r{do}|hTc6r{po}]
          E[hPc2r{u}]O[hPTc2r{od}|hPTc2r{z}|hTc2r{z}]$
vypadat R--s[i1]1(hPTc1)2A[hPTc4r{na}]2V[hPTc7r{s}]J[hPTc1r{jako}]
          $
vytvořit        RSEs[i1]1[hPTc1]2CI[hTc7]EJ[hTc6r{na}|hTc6r{v}]@
vytvořit        R--s[i1]1(hPTc1)2[hTc4]2EJ[hTc6r{na}]%$
vytvořit        R--s[i1]1(hPTc1)2[hTc4]2[hPc3]2EJV(hTc6r{v})%$
vzít    RSEs[i1]1[hPTc1]2[hPTc4r{o}]2S[hPTc4r{za}]2O[hTc2r{z}]
          EJ[hTc6r{na}|hTc6r{v}]V[hPc7r{s}|hTc6r{v}]@
vzít    RSIs[i1]1[hPTc1]2[hPTc4]2CI[hTc7]@
vzít    RSIs[i1]1[hPTc1]2[hTc4]2O(hTc2r{z})2EJV(hTc6r{v})
          A(hTc2r{do}|hTc4r{na})@
vzít    R--s[i1]1(hPTc1)2[hPc4|hPc3]2[hPTc4]2O[hTc2r{z}]
          J[hTc2r{kolem}|hTc4r{za}]%$
vzít    R--s[i1]1(hPTc1)2[hPc4|hPc3]2(hPTc4)2O(hTc2r{z})
          2ACJ(hTc4r{za})%$
vzít    R--s[i1]1(hPTc1)2[hPc4]2CI[hTc7]AJU[hTc4r{přes}|hTc6r{po}]
          %$
vzít    R--s[i1]1(hPTc1)2[hPc4]2O(hTc2r{z})
          A(hTc2r{do}|hTc3r{k}|hTc4r{na}|hTc4r{za})
          J(hTc2r{kolem}|hTc4r{za})%$
vzít    R--s[i1]1(hPTc1)2[hTc4]2O(hTc2r{z})
          A(hPTc4r{na}|hTc2r{do}|hTc3r{k}|hTc4r{za})%$
vzít    R--s[i1]1(hPTc1)2O[hTc2r{z}]A[hTc2r{do}|hTc3r{k}|hTc4r{na}]
          EJU[hTc2r{kolem}]$
vzniknout       R--s[i1]1(hPTc1)2O[hTc2r{z}]EJ[hTc6r{na}|hTc6r{v}]$
začínat RSIs[i1]1[hPTc1]2[hTc4]2V(hPTc7r{s})@
začínat R--s[i1]1(hPTc1)2[hTc4]2CI(hTc7)2V(hPTc7r{s})%$
začít   RSIs[i1]1[hPTc1]2[hTc4]2V(hPTc7r{s})@
začít   R--s[i1]1(hPTc1)2[hTc4]2CI(hTc7)2V(hPTc7r{s})%$
zahájit R--s[i1]1(hPTc1)2[hTc4]2AP[hTc4r{na}]%$
zahájit R--s[i1]1(hPTc1)2[hTc4]2CI(hTc7)EJ(hTc6r{na}|hTc6r{v})%$
zaplatit        R--s[i1]1(hPTc1)2[hM]2CI[hTc7]%$
zaplatit        R--s[i1]1(hPTc1)2[hTc4]2[hPc3]2ACJ(hTc4r{za})%$
zdát    RSEs[i1]1[hPTc1]2[hQc7]2IS[hPTc7]@
získat  RSIs[i1]1[hPTc1]2[hPTc4]2CI(hTc7)@
získat  R--s[i1]1(hPTc1)2EJ[hTc6r{na}]$
získat  R--s[i1]1(hPTc1)2[hPTc4]2CI(hTc7)2S(hPTc4r{za})
          AP(hTc3r{k}|hTc4r{na})%$
získat  R--s[i1]1(hPTc1)2[hPTc4]2S(hPTc4r{za})
          AP(hTc3r{k}|hTc4r{na})%$
```

```
získat   R--s[i1]1(hPTc1)2[hTc4]2E[hPTc6r{na}]O[hPc2r{od}|hTc2r{z}]
            %$
získat   R--s[i1]1(hPTc1)2[hTc4]2[hPc3|hPc4r{pro}]2E[hPTc6r{na}]
            O[hPc2r{od}|hTc2r{z}]%$
změnit   R--s[i1]1(hPTc1)2[hPTc4]2AP(hTc3r{k})%$
znamenat        RSIs[i1]1[hPTc1]2[hPTc4]2A(hTc2r{do})2EJV(hTc6r{v})@
znamenat        R--s[i1]1(hPTc1)2[hPTc4]2CI[hTc7]%$
znát     RSEs[i1]1[hPTc1]2A[hPTc3r{ke}]2V[hPTc7r{s}]@
znát     R--s[i1]1(hPTc1)2[hTc4]2E[hPTc6r{na}]%$
zůstat   R--s[i1]1(hPTc1)2AP(hTc4r{na})E[hPc2r{u}]$
zůstat   R--s[i1]1(hPTc1)2A[sI]2IS[hPTc7]
            E[hTc2r{u}|hTc6r{na}|hTc6r{při}|hTc6r{v}]
            V[hPc7r{s}|hTc2r{bez}|hTc6r{v}]%$
zůstat   R--s[i1]1(hPTc1)2[hPc3]2AH(hPTc6r{po})%$
zvýšit   RSEs[i1]1[hPTc1]2CI[hTc7]2EJV[hTc6r{v}]F[hMr{o}|hTc4r{o}]
            @
žít      R--s[i1]1(hPTc1)2[hTc4|hPTc3|hPTc4r{pro}|hTc6r{o}]
            2O[hPTc2r{z}]C[hTc7|hTc4r{za}]E[hPc2r{u}|hPTc6r{v}|hTc6r{na}]
            %$
```

E. Classification of Czech frames

E.1. Automatically processed frames

These are frames where all inner participants got only one functor. They were processed fully automatically. There are 1312 such classes and they contain 7429 frames.

We list here classes which contain at least ten verbs. The number in the first column shows the frequency of the frame in the dictionary.

```
1435 R--s[i1]1(hPTc1)2[hTc4]%$
1000 R--s[i1]1(hPTc1)2[hPTc4]%$
348 R--s[i1]1(hPTc1)2[hTc4]3[hPc3]%$
238 R--s[i1]1(hPTc1)2[hTc4]3(hPc3)%$
181 R--s[i1]1(hPTc1)$
159 R--s[i1]1(hPTc1)2[hPc4]%$
140 RSEs[i1]1[hPTc1]@
106 RSEs[i1]1[hPTc1]EJ[hTc6r{na}|hTc6r{v}]@
99 RSIs[i1]1[hPTc1]2[hTc4]@
71 R--s[i1]1(hPTc1)EJ[hTc6r{na}|hTc6r{v}]$
68 R--s[i1]1(hPTc1)2[hTc4]EJ(hTc6r{na}|hTc6r{v})%$
61 R--s[i1]1(hPTc1)2[hPc3]%$
56 R--s[i1]1(hPTc1)2[hTc4]A(hTc2r{do}|hTc4r{na})%$
55 R--s[i1]1(hPTc1)2[hPTc4]EJ(hTc6r{na}|hTc6r{v})%$
52 RSEs[i1]1[hPTc1]2[hTc2]@
51 RSEs[i1]1[hPTc1]A[hTc2r{do}|hTc4r{na}]@
50 R--s[i1]1(hPTc1)2[sD|sR|sZ]%$
48 R--s[i1]1(hPTc1)2[hPTc4]A(hTc2r{do}|hTc4r{na})%$
45 R--s[i1]1(hPTc1)2[hTc2|hTc4]%$
39 RSEs[i1]1[hPTc1]2[hPc3]@
36 R--s[i1]1(hPTc1)2[hTc2]%$
29 R--s[i1]1(hPTc1)2[sD]%$
27 RSIs[i1]1[hPTc1]2[hPTc4]@
27 R--s[i1]1(hPTc1)2[hPTc4]3(hPc3)%$
25 R--s[i1]1(hPTc1)2[hPTc3]%$
23 RSEs[i1]1[hPTc1]2[hPTc2]@
20 R--s[i1]1(hPTc1)A[hTc2r{do}|hTc4r{na}]$
20 R--s[i1]1(hPTc1)2[hPTc4]A(hTc2r{do}|hTc3r{k}|hTc4r{na})%$
```

```
19 RSEs[i1]1[hPTc1]2[hPTc3]@
19 R--s[i1]1(hPTc1)2[hTc4|hPc3]%$
18 RSEs[i1]1[hPTc1]C[hTc7|hTc2r{od}]@
18 R--s[i1]1(hPTc1)2[hPTc4]3[hPc3]%$
17 R--s[i1]1(hPTc1)AJTUY[hTc6r{po}]$
17 R--s[i1]1(hPTc1)2[hTc4]3(hPc3)EJ(hTc6r{na}|hTc6r{v})%$
17 R--s[i1]1(hPTc1)2[hPTc4]AP(hTc3r{k}|hTc4r{na})%$
16 RSIs[i1]1[hPTc1]2[hTc2|hTc4]@
16 R--s[i1]1(hPTc1)2[hPc4]A(hTc2r{do}|hTc4r{na})%$
15 RSIs[i1]1[hPTc1]2[sD|sR|sZ]@
15 RSEs[i1]1[hPTc1]A[hPTc3r{ke}|hTc2r{do}|hTc4r{na}]@
15 R--s[i1]1(hPTc1)2[hTc4]EJ[hTc6r{na}|hTc6r{v}]%$
15 R--s[i1]1(hPTc1)2[hTc4]AP(hTc3r{k}|hTc4r{na})%$
14 RSEs[i1]1[hPTc1]J[hTc6r{na}|hTc6r{po}|hTc6r{v}]@
13 R--s[i1]1(hPTc1)2[hTc2|hTc4]3[hPc3]%$
13 R--s[i1]1(hPTc1)2[hPTc4|sD|sR|sZ]%$
12 RSIs[i1]1[hPTc1]2[hTc4]EJ(hTc6r{na}|hTc6r{v})@
12 RSIs[i1]1[hPTc1]2[hPc4]@
12 RSEs[i1]1[hPTc1]E[hPc2r{u}|hTc6r{na}|hTc6r{v}]@
12 RSEs[i1]1[hPTc1]AP[hTc3r{k}|hTc4r{na}]@
12 RSEs[i1]1[hPTc1]2[sD|sR|sZ]@
12 R--s[i1]1(hPTc1)A[hPTc4r{na}|hTc2r{do}]$
12 R--s[i1]1(hPTc1)2[hTc2|hPTc4]%$
12 R--s[i1]1(hPTc1)2[hPTc2]%$
11 RSEs[i1]1[hPTc1]A[hTc2r{do}|hTc3r{k}]@
11 R--s[i1]1(hPTc1)E[hPc2r{u}|hTc6r{na}|hTc6r{v}]$
11 R--s[i1]1(hPTc1)A[hPTc3r{ke}|hTc2r{do}|hTc4r{na}]$
11 R--s[i1]1(hPTc1)2[hTc4]A(hTc2r{do}|hTc3r{k})%$
10 RSIs[i1]1[hPTc1]@
10 R--s[i1]1(hPTc1)V[hTc6r{v}|hTc7r{s}]$
10 R--s[i1]1(hPTc1)J[hTc6r{na}|hTc6r{po}]$
10 R--s[i1]1(hPTc1)2[hTc4]O(hTc2r{od}|hTc2r{z})%$
10 R--s[i1]1(hPTc1)2[hTc4]A(hTc2r{do}|hTc3r{k}|hTc4r{na})%$
10 R--s[i1]1(hPTc1)2[hPTc4]A(hTc2r{do}|hTc4r{v})%$
10 R--s[i1]1(hPTc1)2[hPTc4]A(hTc2r{do}|hTc3r{k})%$
```

E.2. Ambiguous frames

These are frames where some participants are ambiguous between an actant and a free modification. There are 2666 ambiguous classes and they contain 11200 frames.

```
685 R--s[i1]1(hPTc1)2[hPTc4]2CM(hTc7)%$
442 R--s[i1]1(hPTc1)2[hTc4]2CM(hTc7)%$
294 RSEs[i1]1[hPTc1]2CM[hTc7]@
284 R--s[i1]1(hPTc1)2[hPc4]2CM(hTc7)%$
```

```
170 R--s[i1]1(hPTc1)2CM[hTc7]%$
169 R--s[i1]1(hPTc1)2[hTc4]2O(hTc2r{z})%$
131 RSEs[i1]1[hPTc1]2EJV[hTc6r{v}]@
125 R--s[i1]1(hPTc1)2[hTc4]2EJV(hTc6r{v})%$
120 RSEs[i1]1[hPTc1]2CM[hTc7]2EJV[hTc6r{v}]@
119 R--s[i1]1(hPTc1)2[hTc4]2A(hTc2r{do})%$
98 R--s[i1]1(hPTc1)2[hTc4]2(hPc3)2CM(hTc7)%$
97 R--s[i1]1(hPTc1)2[hPTc4]2CM(hTc7)2EJV(hTc6r{v})%$
94 R--s[i1]1(hPTc1)2EJV[hTc6r{v}]$
83 R--s[i1]1(hPTc1)2[hPTc4]2EJV(hTc6r{v})%$
82 R--s[i1]1(hPTc1)2[hPTc4]2A(hTc2r{do})%$
78 RSEs[i1]1[hPTc1]2O[hTc2r{z}]@
73 R--s[i1]1(hPTc1)2[hTc4]2AP(hTc4r{na})%$
66 RSEs[i1]1[hPTc1]2A[hTc2r{do}]@
64 R--s[i1]1(hPTc1)2A[hPTc4r{na}]$
57 R--s[i1]1(hPTc1)2[hTc4]2CM[hTc7]%$
55 RSEs[i1]1[hPTc1]2V[hPTc7r{s}]@
54 R--s[i1]1(hPTc1)2[hTc4]2CM(hTc7)2EJV(hTc6r{v})%$
54 R--s[i1]1(hPTc1)2CM[hTc7]2EJV[hTc6r{v}]%$
53 R--s[i1]1(hPTc1)2[hPTc4]2O(hTc2r{z})%$
48 RSEs[i1]1[hPTc1]2CM[hTc7]EJ[hTc6r{na}|hTc6r{v}]@
47 RSEs[i1]1[hPTc1]2CM[hTc7]2O[hTc2r{z}]@
47 R--s[i1]1(hPTc1)2[hTc4]2V(hPc7r{s})%$
43 R--s[i1]1(hPTc1)2AP[hTc4r{na}]$
41 RSEs[i1]1[hPTc1]2[hPc3]2CM(hTc7)@
41 RSEs[i1]1[hPTc1]2V[hPc7r{s}]@
41 R--s[i1]1(hPTc1)2[hTc4]2(hPc3)2O(hTc2r{z})%$
38 R--s[i1]1(hPTc1)2[hPTc4]2AP(hTc4r{na})%$
38 R--s[i1]1(hPTc1)2[hPc3]2EJV(hTc6r{v})%$
37 R--s[i1]1(hPTc1)2[hPc3]2CM(hTc7)%$
35 RSEs[i1]1[hPTc1]2AP[hTc4r{na}]@
35 R--s[i1]1(hPTc1)2[hTc4]2AP(hTc3r{k})%$
35 R--s[i1]1(hPTc1)2[hPTc4]2O[hTc2r{z}]%$
34 R--s[i1]1(hPTc1)2[hTc4]2(hPc3)2EJV(hTc6r{v})%$
33 R--s[i1]1(hPTc1)2[hTc4]2AP[hTc4r{na}]%$
33 R--s[i1]1(hPTc1)2[hTc4]2A(hTc2r{do})2O(hTc2r{z})%$
33 R--s[i1]1(hPTc1)2[hTc4]2(hPc3)2A(hTc2r{do})%$
33 R--s[i1]1(hPTc1)2[hPTc4]2CM[hTc7]%$
32 R--s[i1]1(hPTc1)2[hPTc4]2CM(hTc7)2AP(hTc4r{na})%$
32 R--s[i1]1(hPTc1)2[hPc4]2A(hTc2r{do})%$
32 R--s[i1]1(hPTc1)2CM[hTc7]2O[hTc2r{z}]%$
32 R--s[i1]1(hPTc1)2A[hTc2r{do}]$
31 RSEs[i1]1[hPTc1]2A[hTc2r{do}]2EJV[hTc6r{v}]@
31 RSEs[i1]1[hPTc1]2A[hPTc4r{na}]@
31 R--s[i1]1(hPTc1)2[hTc4]2CM(hTc7)2A(hTc2r{do})%$
31 R--s[i1]1(hPTc1)2O[hTc2r{z}]$
```

```
30 R--s[i1]1(hPTc1)2[hTc4]2[hPc3]2CM(hTc7)%$
30 R--s[i1]1(hPTc1)2[hPTc4]2CM(hTc7)2A(hTc2r{do})%$
30 R--s[i1]1(hPTc1)2[hPc4]2O(hTc2r{z})%$
29 R--s[i1]1(hPTc1)2[hTc4]2EJ(hTc6r{na})%$
29 R--s[i1]1(hPTc1)2[hTc4]2A(hTc2r{do})2EJV(hTc6r{v})%$
28 RSIs[i1]1[hPTc1]2[hTc4]2CM(hTc7)@
28 R--s[i1]1(hPTc1)2[hTc4]2B(hPc4r{na})%$
28 R--s[i1]1(hPTc1)2[hTc4]2A[hTc2r{do}]%$
28 R--s[i1]1(hPTc1)2[hPTc4]2AP(hTc3r{k})%$
27 R--s[i1]1(hPTc1)2[hTc4]2[hPc3]2O(hTc2r{z})%$
27 R--s[i1]1(hPTc1)2V[hPTc7r{s}]$
27 R--s[i1]1(hPTc1)2V[hPc7r{s}]2JY(hTc6r{o})$
26 R--s[i1]1(hPTc1)2[hTc4]2(hPc3)2AP(hTc4r{na})%$
26 R--s[i1]1(hPTc1)2B[hPc4r{na}]$
24 R--s[i1]1(hPTc1)2[hTc4]2O[hTc2r{z}]%$
24 R--s[i1]1(hPTc1)2[hPTc4]2O(hTc2r{z})A(hTc2r{do}|hTc4r{na})%$
24 R--s[i1]1(hPTc1)2V[hPc7r{s}]$
24 R--s[i1]1(hPTc1)2MS[hPTc7]%$
23 R--s[i1]1(hPTc1)2AP[hTc3r{k}]$
22 RSEs[i1]1[hPTc1]2[hPc3]2EJV(hTc6r{v})@
22 R--s[i1]1(hPTc1)2[hPc4]2AP(hTc3r{k})%$
21 R--s[i1]1(hPTc1)2[hTc4]2A[hPTc4r{na}]%$
21 R--s[i1]1(hPTc1)2[hPTc4]2A[hTc2r{do}]%$
21 R--s[i1]1(hPTc1)2[hPc4]2EJV(hTc6r{v})%$
21 R--s[i1]1(hPTc1)2[hPc3]2AP(hTc4r{na})%$
20 RSEs[i1]1[hPTc1]2O[hTc2r{z}]A[hTc2r{do}|hTc4r{na}]@
20 RSEs[i1]1[hPTc1]2CM[hTc7]2EJ[hTc6r{na}]@
20 RSEs[i1]1[hPTc1]2AP[hTc3r{k}]@
20 R--s[i1]1(hPTc1)2[hTc4]2O(hTc2r{z})2EJV(hTc6r{v})%$
20 R--s[i1]1(hPTc1)2[hTc4]2O(hTc2r{z})2EJ(hTc6r{na})%$
20 R--s[i1]1(hPTc1)2[hTc4]2EJV[hTc6r{v}]%$
20 R--s[i1]1(hPTc1)2[hTc4]2A(hPTc4r{na})%$
20 R--s[i1]1(hPTc1)2[hPTc4]2CM(hTc7)2O(hTc2r{z})%$
19 RSEs[i1]1[hPTc1]2CM[hTc7]2A[hTc2r{do}]@
19 R--s[i1]1(hPTc1)2[hTc4]2[hPc3]2A(hTc2r{do})%$
19 R--s[i1]1(hPTc1)2[hTc4]2O[hPTc2r{z}]%$
19 R--s[i1]1(hPTc1)2[hPTc4]2A(hTc2r{do})2O(hTc2r{z})%$
19 R--s[i1]1(hPTc1)2[hPc4]2CM[hTc7]%$
19 R--s[i1]1(hPTc1)2MV[hTc7r{s}]$
18 R--s[i1]1(hPTc1)2[hPTc4]2CM(hTc7)EJ(hTc6r{na}|hTc6r{v})%$
18 R--s[i1]1(hPTc1)2[hPc3]2A(hTc2r{do})%$
18 R--s[i1]1(hPTc1)2CM[hTc7]2EJV(hTc6r{v})%$
17 RSEs[i1]1[hPTc1]2EJ[hTc6r{na}]@
17 RSEs[i1]1[hPTc1]2A[hTc2r{do}]2O[hTc2r{z}]@
17 RSEs[i1]1[hPTc1]2A[hPTc3r{ke}]@
17 R--s[i1]1(hPTc1)2[hTc4]2CM(hTc7)2O(hTc2r{z})%$
```

```
17 R--s[i1]1(hPTc1)2[hTc4]2CDOZ(hTc2r{od})%$
17 R--s[i1]1(hPTc1)2[hPTc4]2V[hPTc7r{s}]%$
17 R--s[i1]1(hPTc1)2EJ[hTc6r{na}]$
17 R--s[i1]1(hPTc1)2CM[hTc7]EJ[hTc6r{na}|hTc6r{v}]%$
16 RSEs[i1]1[hPTc1]2O[hTc2r{z}]2AP[hTc4r{na}]@
16 R--s[i1]1(hPTc1)2[hTc4]2[hPc3]2A[hTc2r{do}]%$
16 R--s[i1]1(hPTc1)2[hTc4]2O(hTc2r{z})A(hTc2r{do}|hTc4r{na})%$
16 R--s[i1]1(hPTc1)2[hTc4]2CM(hTc7)2AP(hTc4r{na})%$
16 R--s[i1]1(hPTc1)2[hPTc4]2A(hTc2r{do})2EJV(hTc6r{v})%$
16 R--s[i1]1(hPTc1)2CM[hTc7]2AP(hTc4r{na})%$
15 RSEs[i1]1[hPTc1]2[hPc3]2O[hTc2r{z}]@
15 RSEs[i1]1[hPTc1]2CM[hTc7]2A(hTc2r{do})@
15 R--s[i1]1(hPTc1)2[hTc4]2O(hPTc2r{z})%$
15 R--s[i1]1(hPTc1)2[hPTc3]2CM(hTc7)%$
15 R--s[i1]1(hPTc1)2[hPc4]2ACJ(hTc4r{za})%$
15 R--s[i1]1(hPTc1)2CM[hTc7]2A(hTc2r{do})%$
15 R--s[i1]1(hPTc1)2A[hTc2r{do}]2O[hTc2r{z}]$
14 RSIs[i1]1[hPTc1]2[hTc4]2EJV(hTc6r{v})@
14 R--s[i1]1(hPTc1)2[hTc4]2[hPc3]2O[hTc2r{z}]%$
14 R--s[i1]1(hPTc1)2[hTc4]2CM(hTc7)2EJ(hTc6r{na})%$
14 R--s[i1]1(hPTc1)2[hPTc4]2MS(hPTc7)%$
14 R--s[i1]1(hPTc1)2[hPc3]2O(hTc2r{z})%$
14 R--s[i1]1(hPTc1)2CM[hTc7]2AP[hTc4r{na}]%$
13 RSEs[i1]1[hPTc1]2[hPc3]2CM[hTc7]@
13 RSEs[i1]1[hPTc1]2A[hPTc2r{do}]@
13 R--s[i1]1(hPTc1)2[hPTc4]2CDOZ(hTc2r{od})%$
13 R--s[i1]1(hPTc1)2[hPTc4]2ACJ(hTc4r{za})%$
13 R--s[i1]1(hPTc1)2[hPTc3]2EJV(hTc6r{v})%$
13 R--s[i1]1(hPTc1)2[hPc4]2A[hTc2r{do}]%$
13 R--s[i1]1(hPTc1)2CM[hTc7]2B(hPc4r{na})%$
13 R--s[i1]1(hPTc1)2A[hPTc2r{do}]$
13 R--s[i1]1(hPTc1)2AI[sI]%$
12 RSEs[i1]1[hPTc1]2[hPTc3]2CM(hTc7)2EJV(hTc6r{v})@
12 RSEs[i1]1[hPTc1]2[hPc3]2O(hTc2r{z})@
12 RSEs[i1]1[hPTc1]2O[hTc2r{z}]2EJV[hTc6r{v}]@
12 RSEs[i1]1[hPTc1]2O[hPTc2r{od}]@
12 RSEs[i1]1[hPTc1]2MV[hTc7r{s}]@
12 R--s[i1]1(hPTc1)2[hTc4]2MV[hTc7r{s}]%$
12 R--s[i1]1(hPTc1)2[hPTc4]2V(hPTc7r{s})%$
12 R--s[i1]1(hPTc1)2[hPTc3]2CM(hTc7)2EJV(hTc6r{v})%$
12 R--s[i1]1(hPTc1)2[hPc4]2O[hTc2r{z}]%$
12 R--s[i1]1(hPTc1)2[hPc4]2CM(hTc7)2AP(hTc3r{k})%$
12 R--s[i1]1(hPTc1)2[hPc4]2AF(hTc4r{o})%$
12 R--s[i1]1(hPTc1)2[hPc3]2EJ(hTc6r{na})%$
12 R--s[i1]1(hPTc1)2O[hTc2r{z}]2EJV[hTc6r{v}]$
12 R--s[i1]1(hPTc1)2AH[hPTc6r{po}]$
```

```
11 RSIs[i1]1[hPTc1]2V[hPc7r{s}]2JY(hTc6r{o})@
11 R--s[i1]1(hPTc1)2[hTc4]2MV(hTc7r{s})%$
11 R--s[i1]1(hPTc1)2[hTc4]2(hPc3)2EJ(hTc6r{na})%$
11 R--s[i1]1(hPTc1)2[hTc4]2(hPc3)2CM(hTc7)2EJV(hTc6r{v})%$
11 R--s[i1]1(hPTc1)2[hPc4]2V(hPc7r{s})%$
11 R--s[i1]1(hPTc1)2[hPc4]2CM(hTc7)2EJV(hTc6r{v})%$
11 R--s[i1]1(hPTc1)2[hPc4]2AP[hTc3r{k}]%$
11 R--s[i1]1(hPTc1)2[hPc3]2EJV[hTc6r{v}]%$
11 R--s[i1]1(hPTc1)2[hPc3]2A[hTc2r{do}]%$
11 R--s[i1]1(hPTc1)2O[hTc2r{z}]2AP[hTc4r{na}]$
11 R--s[i1]1(hPTc1)2CM[hTc7]2B[hPc4r{na}]%$
10 RSIs[i1]1[hPTc1]2[hTc4]2O(hTc2r{z})@
10 RSIs[i1]1[hPTc1]2A[hPTc4r{na}]@
10 RSEs[i1]1[hPTc1]2[hPTc3]2EJV(hTc6r{v})@
10 RSEs[i1]1[hPTc1]2V[hPTc7r{s}]EJ(hTc6r{na}|hTc6r{v})@
10 RSEs[i1]1[hPTc1]2V[hPTc7r{s}]2EJV(hTc6r{v})@
10 RSEs[i1]1[hPTc1]2CM[hTc7]2AP[hTc4r{na}]@
10 RSEs[i1]1[hPTc1]2AP[hTc4r{na}]2EJV[hTc6r{v}]@
10 R--s[i1]1(hPTc1)2[hTc4]2O(hPc2r{od})%$
10 R--s[i1]1(hPTc1)2[hTc4]2JY(hTc6r{o})%$
10 R--s[i1]1(hPTc1)2[hTc4]2EJ(hTc6r{na})2EJV(hTc6r{v})%$
10 R--s[i1]1(hPTc1)2[hTc4]2E(hPTc6r{v})%$
10 R--s[i1]1(hPTc1)2[hTc4]2E(hPc7r{před})%$
10 R--s[i1]1(hPTc1)2[hTc4]2B[hPc4r{na}]%$
10 R--s[i1]1(hPTc1)2[hTc4]2A[hPTc2r{do}]%$
10 R--s[i1]1(hPTc1)2[hPTc4]2O(hTc2r{z})2EJV(hTc6r{v})%$
10 R--s[i1]1(hPTc1)2[hPTc4]2O(hPTc2r{od})%$
10 R--s[i1]1(hPTc1)2[hPc3]2AP(hTc3r{k})%$
10 R--s[i1]1(hPTc1)2MS[hPTc7]2V[hPTc7r{s}]%$
10 R--s[i1]1(hPTc1)2AF[hTc4r{o}]$
```

F. Experiment with LFG

F.1. Verb lexicon

```
běhat   V at @(INTRANSREFLPASSKDE běhat).
boj     V bat { @(RFLSENOPASS bát_se) | @(2RFLSENOPASS bát_se)
                | @(IRFLSENOPASSSUBJ bát_se) }.
dokázat V at @(IRFLPASSSUBJ dokázat).
dopátrat        V at @(2RFLSENOPASS dopátrat_se).
dychtit V it @(IRFLPASSSUBJ dychtit).
hatit   V it @(TRANSRFLPERPASSDAT hatit).
hnízdit V it @(INTRANSREFLPASSKDE hnízdit).
chápat  V at @(SRFLPERPASSZE chápat).
chránit V it @(2RFLSENOPASS chránit_se).
chvátat V at @(INTRANSREFLPASSKAM chvátat).
kazit   V it @(RFLSENOPASS kazit_se).
konfiskovat     V ovat { @(TRANSRFLPERPASSDAT konfiskovat)
                | @(TRANSRFLPERPASS konfiskovat) }.
křížit  V it @(TRANSRFLPERPASSDAT křížit).
lovit   V it { @(TRANSRFLPERPASS lovit) | @(INTRANSREFLPASS lovit) }.
muset   V et @(MODALRAISESUBJ muset).
nabaštit        V it @(2RFLSENOPASS nabaštit_se).
nacvičit        V it @(RFLSINOPASSACC nacvičit_si).
nachytat        V at @(2RFLPERPASS nachytat).
nalovit V it @(2RFLPERPASS nalovit).
nastěhovat      V ovat @(RFLSENOPASSKAM nastěhovat_se).
nažrat  V at @(2RFLSENOPASS nažrat_se).
plánovat        V ovat @(RFLSINOPASSACC plánovat_si).
popírat V at @(SRFLPERPASSZE popírat).
poputovat       V ovat @(INTRANSREFLPASSKAM poputovat).
pršet   V et @(METEOVERB pršet).
přikazovat      V ovat @(IRFLPERPASSOBJ přikazovat).
přinutit        V it @(IRFLPERPASSOBJ přinutit).
relaxovat       V ovat @(INTRANSREFLPASSKDE relaxovat).
schovávat       V at @(RFLSENOPASSKDE schovávat_se).
vodit   V it { @(TRANSRFLPERPASSKAM vodit) | @(TRANSRFLPERPASS vodit) }.
vyhazovat       V ovat { @(TRANSRFLPERPASSKAM vyhazovat)
                | @(TRANSRFLPERPASS vyhazovat) }.
```

```
vyšplhat        V at @(RFLSENOPASSKAM vyšplhat_se).
zabydlet        V et @(RFLSENOPASSKDE zabydlet_se).
zapomínat       V at @(IRFLPASSSUBJ zapomínat).
```

F.2. Templates

```
TRANSRFLPERPASS(P) =
        @(LR-TRANSRFLPERPASS (^ PRED)='P<(^ SUBJ)(^ OBJ)>').
TRANSRFLPERPASSDAT(P) =
        @(LR-TRANSRFLPERPASS {(^ PRED)='P<(^ SUBJ)(^ OBJ)(^ OBJ2)>'
                        (^ OBJ2 CASE)=DAT}).
TRANSRFLPERPASSKAM(P) =
        @(LR-TRANSRFLPERPASS {(^ PRED)='P<(^ SUBJ)(^ OBJ)(^ ACOMP)>'
                        (^ ACOMP SEM)=KAM}).

INTRANSREFLPASS(P) =
        @(LR-INTRANSREFLPASS (^ PRED)='P<(^ SUBJ)>').
INTRANSREFLPASSKAM(P) =
        @(LR-INTRANSREFLPASS {(^ PRED)='P<(^ SUBJ)(^ ACOMP)>'
                        (^ ACOMP SEM)=KAM}).
INTRANSREFLPASSKDE(P) =
        @(LR-INTRANSREFLPASS {(^ PRED)='P<(^ SUBJ)(^ ACOMP)>'
                        (^ ACOMP SEM)=KDE}).

RFLSENOPASS(P) =
        @(LR-RFLSENOPASS (^ PRED)='P<(^ SUBJ)>').
RFLSENOPASSKAM(P) =
        @(LR-RFLSENOPASS {(^ PRED)='P<(^ SUBJ)(^ ACOMP)>'
                        (^ ACOMP SEM)=KAM}).
RFLSENOPASSKDE(P) =
        @(LR-RFLSENOPASS {(^ PRED)='P<(^ SUBJ)(^ ACOMP)>'
                        (^ ACOMP SEM)=KDE}).

RFLSINOPASSACC(P) =
        @(LR-RFLSINOPASS {(^ PRED)='P<(^ SUBJ)(^ OBJ)>'
                        (^ OBJ CASE)=ACC}).

2RFLSENOPASS(P) =
        @(LR-RFLSENOPASS {(^ PRED)='P<(^ SUBJ)(^ OBJ)>'
                        (^ OBJ CASE)=GEN}).
2RFLPERPASS(P) =
        @(LR-RFLPERPASS {(^ PRED)='P<(^ SUBJ)(^ OBJ)>'
                        (^ OBJ CASE)=GEN}).

IRFLPERPASSOBJ(P) =
```

```
        @(LR-TRANSRFLPERPASS {(^ PRED)='P<(^ SUBJ)(^ OBJ)(^ XCOMP)>'
                        (^ XCOMP SUBJ)=(^ OBJ)
                        (^ XCOMP INF)=c +}).
IRFLPASSSUBJ(P) =
        @(LR-RFLPASS {(^ PRED)='P<(^ SUBJ)(^ XCOMP)>'
                        (^ XCOMP SUBJ)=(^ SUBJ)
                        (^ XCOMP INF)=c +}).

IRFLSENOPASSSUBJ(P) =
        @(LR-RFLSENOPASS {(^ PRED)='P<(^ SUBJ)(^ XCOMP)>'
                        (^ XCOMP SUBJ)=(^ SUBJ)
                        (^ XCOMP INF)=c +}).

SRFLPERPASSZE(P) =
        @(LR-RFLPERPASS {(^ PRED)='P<(^ SUBJ)(^ SCOMP)>'
                        (^ SCOMP CONJ)=že}).

MODALRAISESUBJ(P) =
        { (^ PRED)='P<(^ XCOMP)>(^ SUBJ)'
                (^ XCOMP SUBJ)=(^ SUBJ)
        | (^ PRED)='P<(^ XCOMP)>'
                ~(^ XCOMP SUBJ) }
        (^ XCOMP INF)=c +
        ~(^ REFL).

METEOVERB(P) =
        { (^ PRED)='P<NULL>'
        | (^ PRED)='P<NULL (^ ACOMP)>'
                (^ ACOMP SEM)=KDE}
        ~(^ REFL).
```

F.3. Lexical rules

```
LR-TRANSRFLPERPASS(SCHEMATA) =
        { SCHEMATA
        (^ OBJ CASE)=ACC
        ~(^ REFL)
        |SCHEMATA
        (^ REFL)=c SE
        (^ OBJ)->(^ SUBJ)
        (^ OBJ CASE)=NOM
        (^ SUBJ)->NULL
        |SCHEMATA
        (^ OBJ)->(^ SUBJ)
```

```
        (^ OBJ CASE)=NOM
        (^ SUBJ)->NULL
        ~(^ REFL)
        ~(^ TENSE)
        (^ PARTICIPLE)=c PASS }.

LR-INTRANSREFLPASS(SCHEMATA) =
        { SCHEMATA
        ~(^ REFL)
        |SCHEMATA
        (^ REFL)=c SE
        (^ SUBJ)->NULL
        (^ GENDER)=Neut
        (^ NUM)=SG
        (^ PERSON)=3 }.

LR-RFLPERPASS(SCHEMATA) =
        { SCHEMATA
        ~(^ REFL)
        | { SCHEMATA
        (^ REFL)=c SE
        (^ SUBJ)->NULL
        |SCHEMATA
        (^ SUBJ)->NULL
        ~(^ REFL)
        ~(^ TENSE)
        (^ PARTICIPLE)=c PASS }
        (^ GENDER)=Neut
        (^ NUM)=SG
        (^ PERSON)=3 }.

LR-RFLPASS(SCHEMATA) =
        { SCHEMATA
        ~(^ REFL)
        |SCHEMATA
        (^ REFL)=c SE
        (^ SUBJ)->NULL
        (^ GENDER)=Neut
        (^ NUM)=SG
        (^ PERSON)=3 }.

LR-RFLSENOPASS(SCHEMATA) =
        SCHEMATA
        (^ REFL)=c SE .

LR-RFLSINOPASS(SCHEMATA) =
```

```
SCHEMATA
(^ REFL)=c SI .
```

F.4. Grammar

```
TEST  CZECH
   NP  → A*:  (↑ ATTR)=↓;
         N.

   PP  → PREP
         NP:  (↑ OBJ)=↓.

   S  →  {  {  NP:  (↑ SUBJ)=↓
                    (↑ SUBJ CASE)=NOM
              |VP:  (↑ SUBJ)=↓
                    (↓ INF)=c +}
            { ADV: ↓∈(↑ ADJ)
             |PP: ↓∈(↑ ADJ)}*
            VP:  (↑ PERSON)=(↑ SUBJ PERSON)
                 (↑ NUM)=(↑ SUBJ NUM)
                 (↑ GENDER)=(↑ SUBJ GENDER)
                 (↑ TENSE)
          |{ ADV:  {  (↑ ACOMP)=↓
                     |↓∈(↑ ADJ)}
             |PP:  {  (↑ ACOMP)=↓
                     |↓∈(↑ ADJ)}}*
            VP:  (↑ PERSON)=3
                 (↑ NUM)=SG
                 (↑ GENDER)=Neut
                 (↑ TENSE)}.

   S'  → CONJ
         S.

   VP  → (PRTCL)
         V
         {  (NP:  (↑ OBJ)=↓)
           |(NP:  (↑ OBJ2)=↓)
           |ADV:  {  (↑ ACOMP)=↓
                    |↓∈(↑ ADJ)}
           |PP:  {  (↑ (↓ PCASE))=(↓ OBJ)
                   |(↑ ACOMP)=↓
                   |↓∈(↑ ADJ)}
           |VP:  (↑ XCOMP)=↓
           |S':  (↑ SCOMP)=↓}*.

----
```

Figure F.1.: Simple grammar in LFG

F.5. Test sentences

the following picture shows sentences used for testing the lexicon. The sentences contain only limited vocabulary because a larger vocabulary would also need a large morphological module. As the main point was to show verb frames we consider this small lexicon sufficient.

The numbers in parentheses mean number of analyses, time of processing and number of steps needed for analysis.

```
S: kočky běhají na zahradě (1 0.27 13)
S: na zahradě se běhá (1 0.15 11)
S: kočky se bojí (1 0.08 8)
S: kočky se bojí na zahradě (1 0.31 15)
S: kočky se bojí v noci (1 0.55 15)
S: kočky se bojí Hanky (1 0.24 12)
S: kočky se bojí lovit myšky (1 1.1 18)
S: kočky se bojí lovit v noci (2 3.36 21)
S: kočky dychtí lovit myšky (1 0.25 14)
S: Hanka hatí kočce lov (1 0.13 14)
S: lov se hatí kočce (1 0.13 12)
S: myšky hnízdí v gauči (1 0.2 13)
S: maso se kazí (1 0.05 8)
S: kočka konfiskuje myšce maso (1 0.13 14)
S: maso se konfiskuje myšce (1 0.16 12)
S: kočky musí lovit myšky (1 0.15 16)
S: myšky se musí lovit (1 0.17 12)
S: musí pršet (1 0.06 10)
S: kočky se nabaští masa (1 0.09 12)
S: kočky naloví myšek (1 0.08 10)
S: tady se naloví myšek  (1 0.14 11)
S: kočky zapomínají lovit myšky (1 0.23 16)
S: myšky se zapomínají lovit (1 0.22 11)
S: Hanka chápe že kočky loví (1 0.18 16)
S: Hanka chápe že myšky se loví v noci (2 1.94 30)
S: Hanka chápe že myšky se musí lovit (1 0.72 27)
S: Hanka přinutí myšky chápat že kočky loví (1 0.58 27)
S: kočky se přinutí lovit myšky (1 1.03 18)
S: kočky se vyšplhají na gauč (1 0.41 15)
```

Figure F.2.: Testing sentences

The pictures on the following pages show c-structures and f-structures of these sentences:

(140) a. *Kočky se bojí lovit myšky.*
Cats$_{Nom}$ SE fear hunt$_{Inf}$ mice$_{Acc}$.

b. *Myšky se zapomínají lovit.*
Mice$_{Nom}$ SE forget hunt$_{Inf}$.
'It's beeing forgotten to hunt mice.'

c. *Hanka chápe že myšky se musí lovit.*
Hanka$_{Nom}$ understands that mice$_{Nom}$ SE must hunt$_{Inf}$.
'Hanka understands that mice must be hunted.'

d. *Hanka přinutí myšky chápat že kočky loví.*
Hanka$_{Nom}$ makes mice$_{Acc}$ understand$_{Inf}$ that cats$_{Nom}$ hunt.

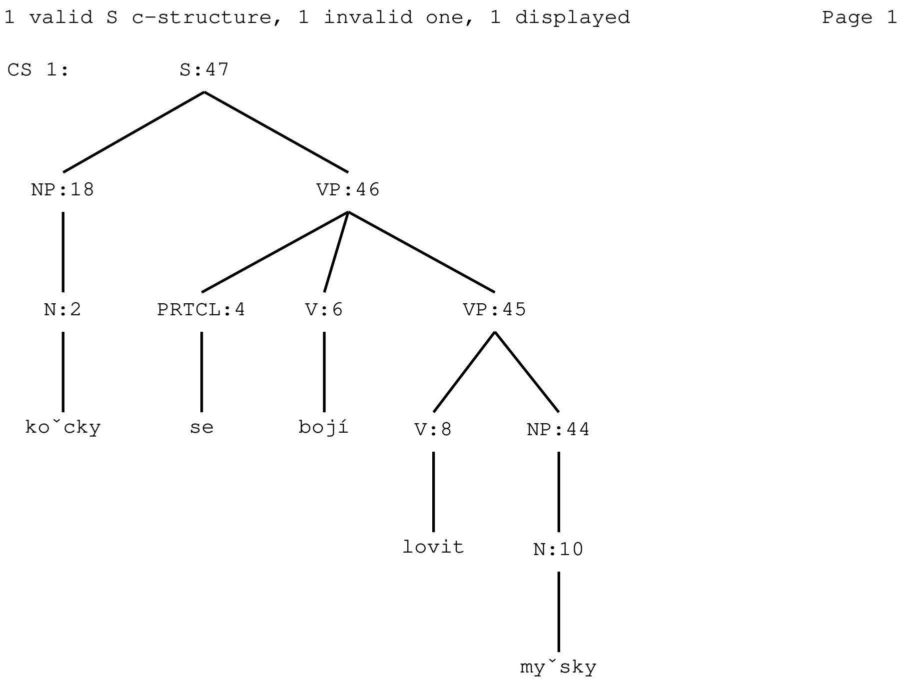


Figure F.3.: C-structure of sentence 140a

```
F-structures for S 47 in CS 1:  1 displayed              Page 1


4 solutions:  2 consistent, 1 complete, 4 coherent
F-structure 1:
   [PRED    'bát_se<[18:kočka], [45:lovit]>'                     ]
   [SUBJ  18[PRED 'kočka'                                ]       ]
   [       2[CASE NOM, GENDER Fem, NUM PL, PERSON 3      ]       ]
   [         [PRED  'lovit<[18:kočka], [44:myška]>'       ]      ]
   [         [INF   +                                     ]      ]
   [         [SUBJ  [18:kočka]                            ]      ]
   [XCOMP    [          [PRED    'myška' ]                ]      ]
   [         [          [CASE    ACC     ]                ]      ]
 47[         [OBJ     44[GENDER  Fem     ]                ]      ]
 46[       45[        10[NUM     PL      ]                ]      ]
  6[        8[                                            ]      ]
  4[GENDER Fem, NUM PL, PERSON 3, REFL SE, TENSE PRES             ]
```

Figure F.4.: F-structure of sentence 140a

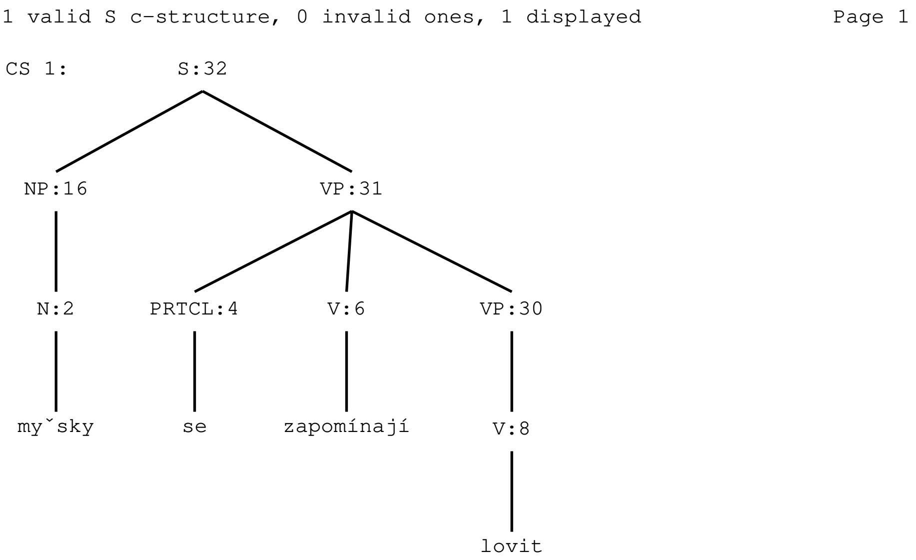


Figure F.5.: C-structure of sentence 140b

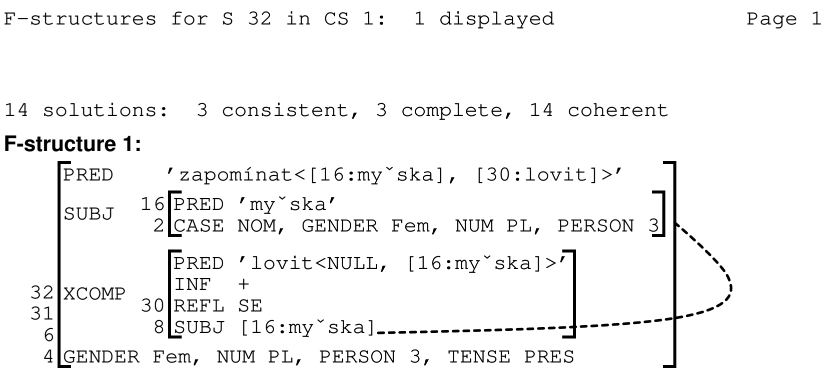


Figure F.6.: F-structure of sentence 140b

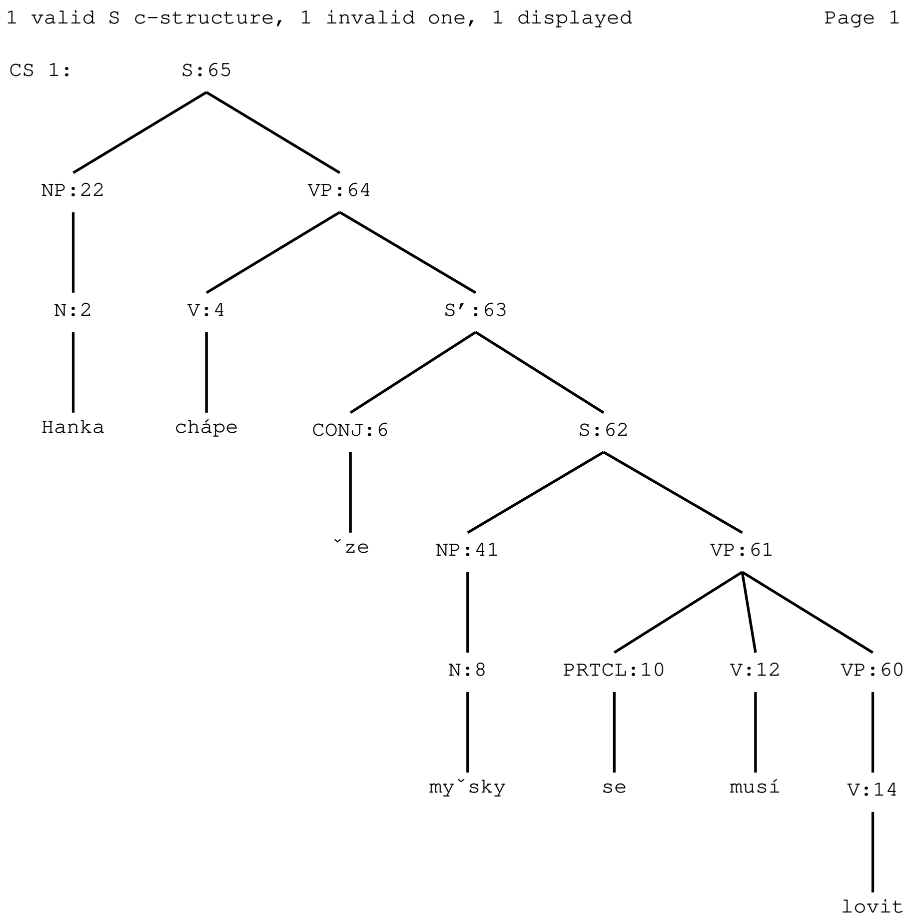


Figure F.7.: C-structure of sentence 140c

```
F-structures for S 65 in CS 1:  1 displayed              Page 1


12 solutions:  2 consistent, 3 complete, 12 coherent
```

F-structure 1:

```
     PRED     'chápat<[22:Hanka], [63:muset]>'
     SUBJ   22 PRED 'Hanka'
             2 CASE NOM, GENDER Fem, NUM SG, PERSON 3
               PRED    'muset<[60:lovit]>[41:myˇska]'
               SUBJ   41 PRED 'myˇska'
                       8 CASE NOM, GENDER Fem, NUM PL, PERSON 3
            63
     SCOMP  62            PRED 'lovit<NULL, [41:myˇska]>'
            61 XCOMP      INF  +
            12         60 REFL SE
   65       10         14 SUBJ [41:myˇska]
   64        6 CONJ ZE, GENDER Fem, NUM PL, PERSON 3, TENSE PRES
    4 GENDER Fem, NUM SG, PERSON 3, TENSE PRES
```

Figure F.8.: F-structure of sentence 140c

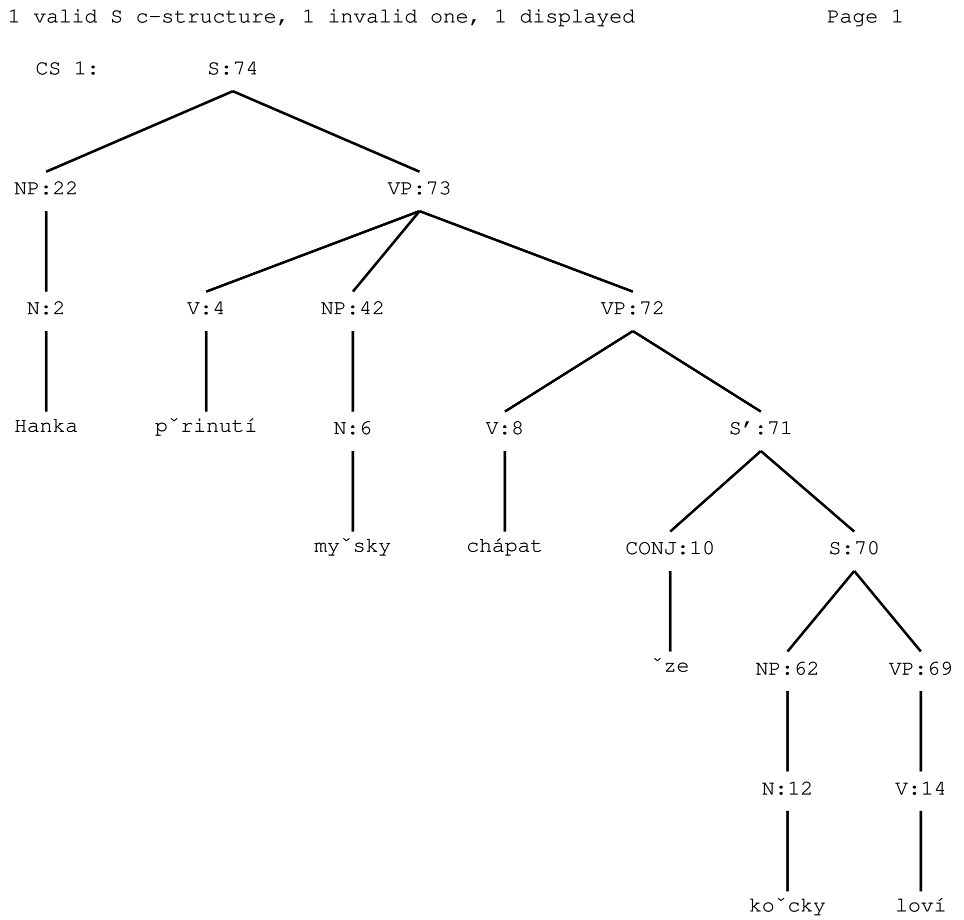


Figure F.9.: C-structure of sentence 140d

```
F-structures for S 74 in CS 1:  1 displayed             Page 1


20 solutions:  20 consistent, 1 complete, 20 coherent
```

F-structure 1:

```
   [PRED     'přinutit<[22:Hanka], [42:myška], [72:chápat]>'                  ]
   [SUBJ   22[PRED 'Hanka'                                  ]                  ]
   [        2[CASE NOM, GENDER Fem, NUM SG, PERSON 3        ]                  ]
   [         [PRED   'myška']                                                  ]
   [         [CASE   ACC    ]                                                  ]
   [OBJ    42[GENDER Fem    ]                                                  ]
   [        6[NUM    PL     ]                                                  ]
   [         [PRED    'chápat<[42:myška], [71:lovit]>'                        ]]
   [         [INF     +                                                       ]]
   [         [SUBJ    [42:myška]                                              ]]
   [XCOMP    [      71[PRED   'lovit<[62:kočka]>'                            ]]]
   [         [      70[                                                      ]]]
   [         [SCOMP 69[SUBJ  62[PRED 'kočka'                             ]   ]]]
   [         [      14[      12[CASE NOM, GENDER Fem, NUM PL, PERSON 3   ]   ]]]
 74[       72[      10[CONJ ZE, GENDER Fem, NUM PL, PERSON 3, TENSE PRES   ]]]
 73[        8[                                                              ]]
  4[GENDER Fem, NUM SG, PERSON 3, TENSE PRES                                  ]
```

Figure F.10.: F-structure of sentence 140d

G. Web interface to the lexicon

The lexicon can be viewed through a web browser. The interface enables various views after many criteria.

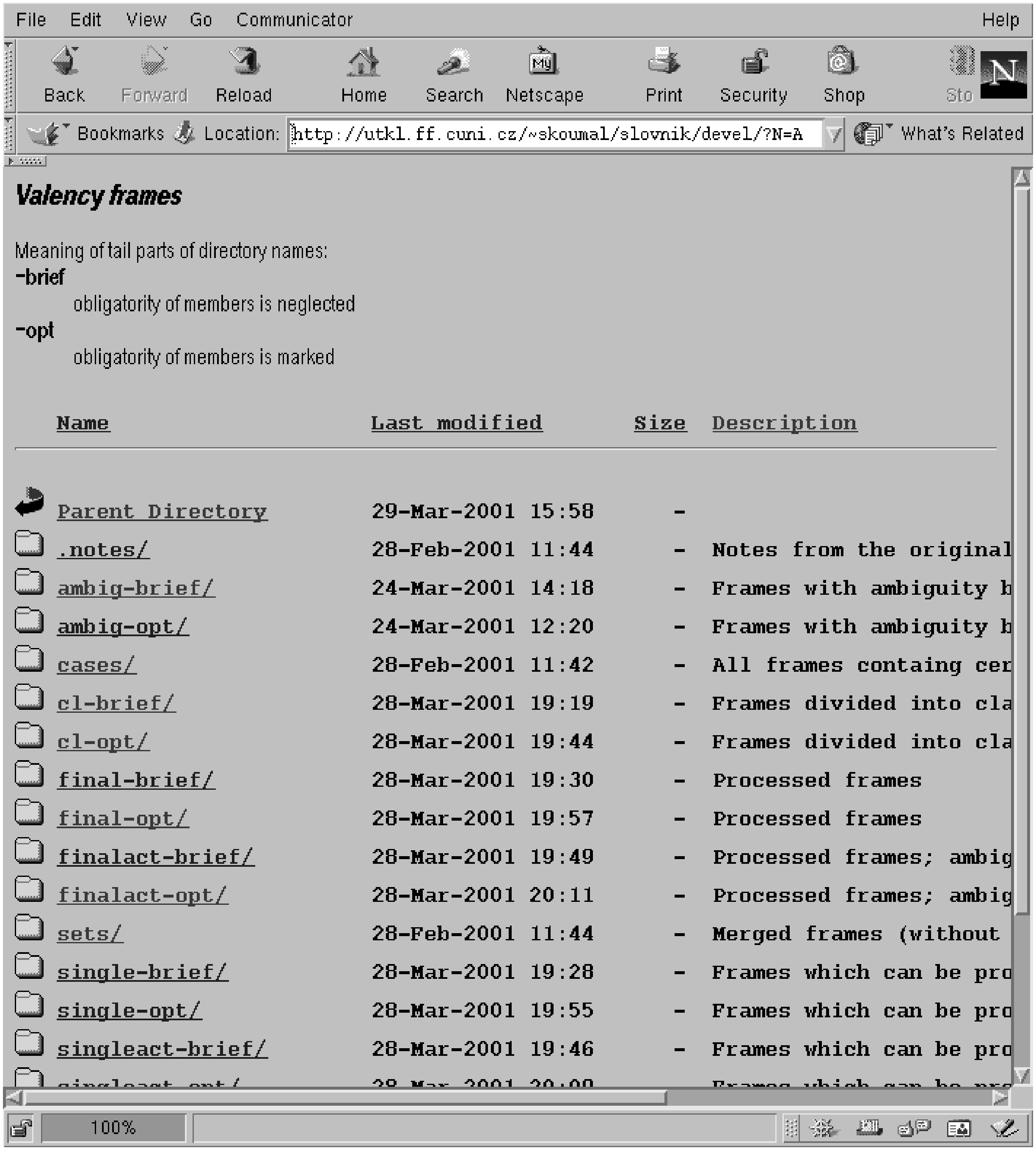
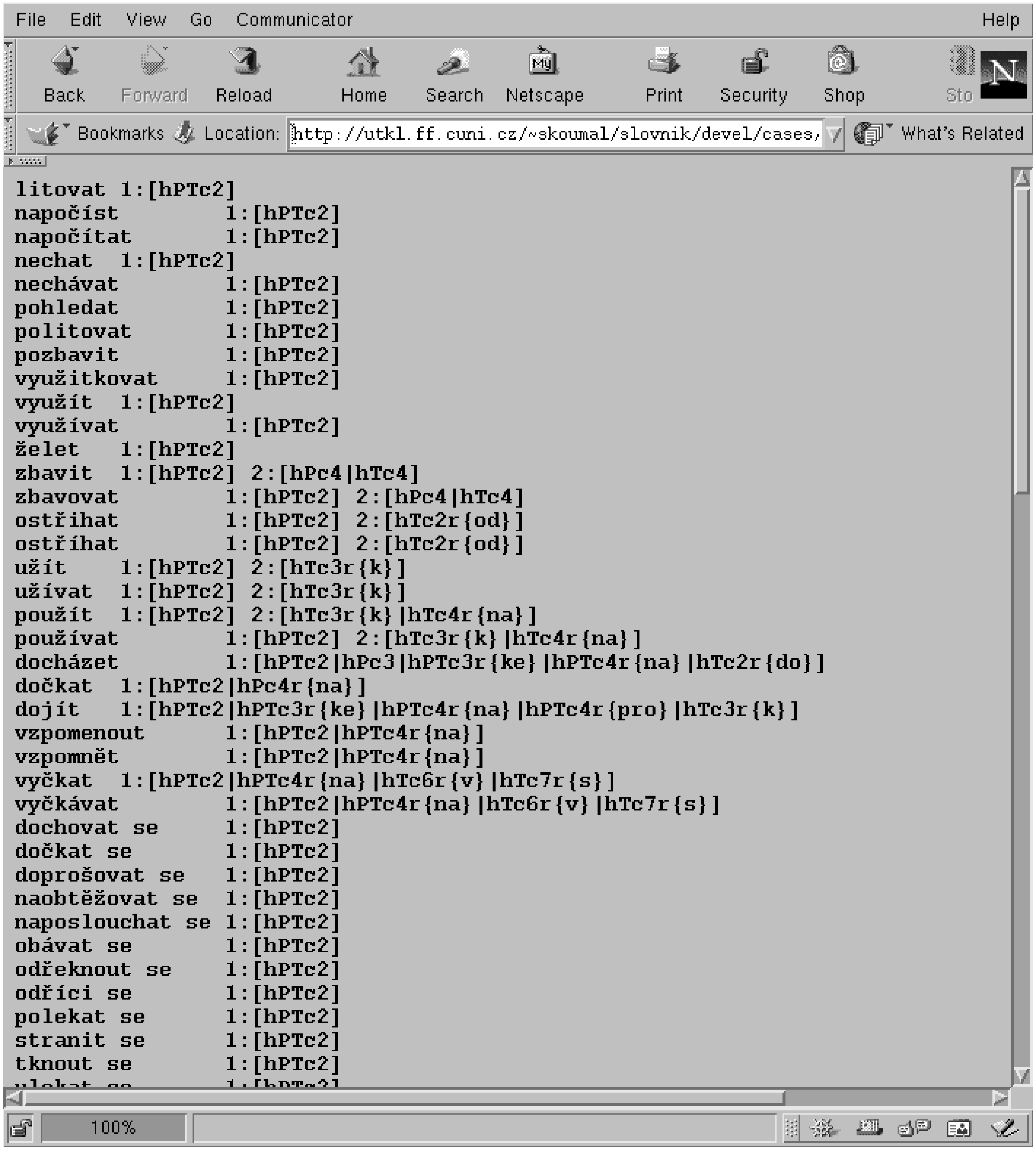


Figure G.1.: Main window of the web interface




```
litovat 1:[hPTc2]
napočíst        1:[hPTc2]
napočítat       1:[hPTc2]
nechat  1:[hPTc2]
nechávat        1:[hPTc2]
pohledat        1:[hPTc2]
politovat       1:[hPTc2]
pozbavit        1:[hPTc2]
využitkovat     1:[hPTc2]
využít  1:[hPTc2]
využívat        1:[hPTc2]
želet   1:[hPTc2]
zbavit  1:[hPTc2] 2:[hPc4|hTc4]
zbavovat        1:[hPTc2] 2:[hPc4|hTc4]
ostřihat        1:[hPTc2] 2:[hTc2r{od}]
ostříhat        1:[hPTc2] 2:[hTc2r{od}]
užít    1:[hPTc2] 2:[hTc3r{k}]
užívat  1:[hPTc2] 2:[hTc3r{k}]
použít  1:[hPTc2] 2:[hTc3r{k}|hTc4r{na}]
používat        1:[hPTc2] 2:[hTc3r{k}|hTc4r{na}]
docházet        1:[hPTc2|hPc3|hPTc3r{ke}|hPTc4r{na}|hTc2r{do}]
dočkat  1:[hPTc2|hPc4r{na}]
dojít   1:[hPTc2|hPTc3r{ke}|hPTc4r{na}|hPTc4r{pro}|hTc3r{k}]
vzpomenout      1:[hPTc2|hPTc4r{na}]
vzpomnět        1:[hPTc2|hPTc4r{na}]
vyčkat  1:[hPTc2|hPTc4r{na}|hTc6r{v}|hTc7r{s}]
vyčkávat        1:[hPTc2|hPTc4r{na}|hTc6r{v}|hTc7r{s}]
dochovat se     1:[hPTc2]
dočkat se       1:[hPTc2]
doprošovat se   1:[hPTc2]
naobtěžovat se  1:[hPTc2]
naposlouchat se 1:[hPTc2]
obávat se       1:[hPTc2]
odřeknout se    1:[hPTc2]
odříci se       1:[hPTc2]
polekat se      1:[hPTc2]
stranit se      1:[hPTc2]
tknout se       1:[hPTc2]
```

Figure G.2.: File with all frames containing hPTc2

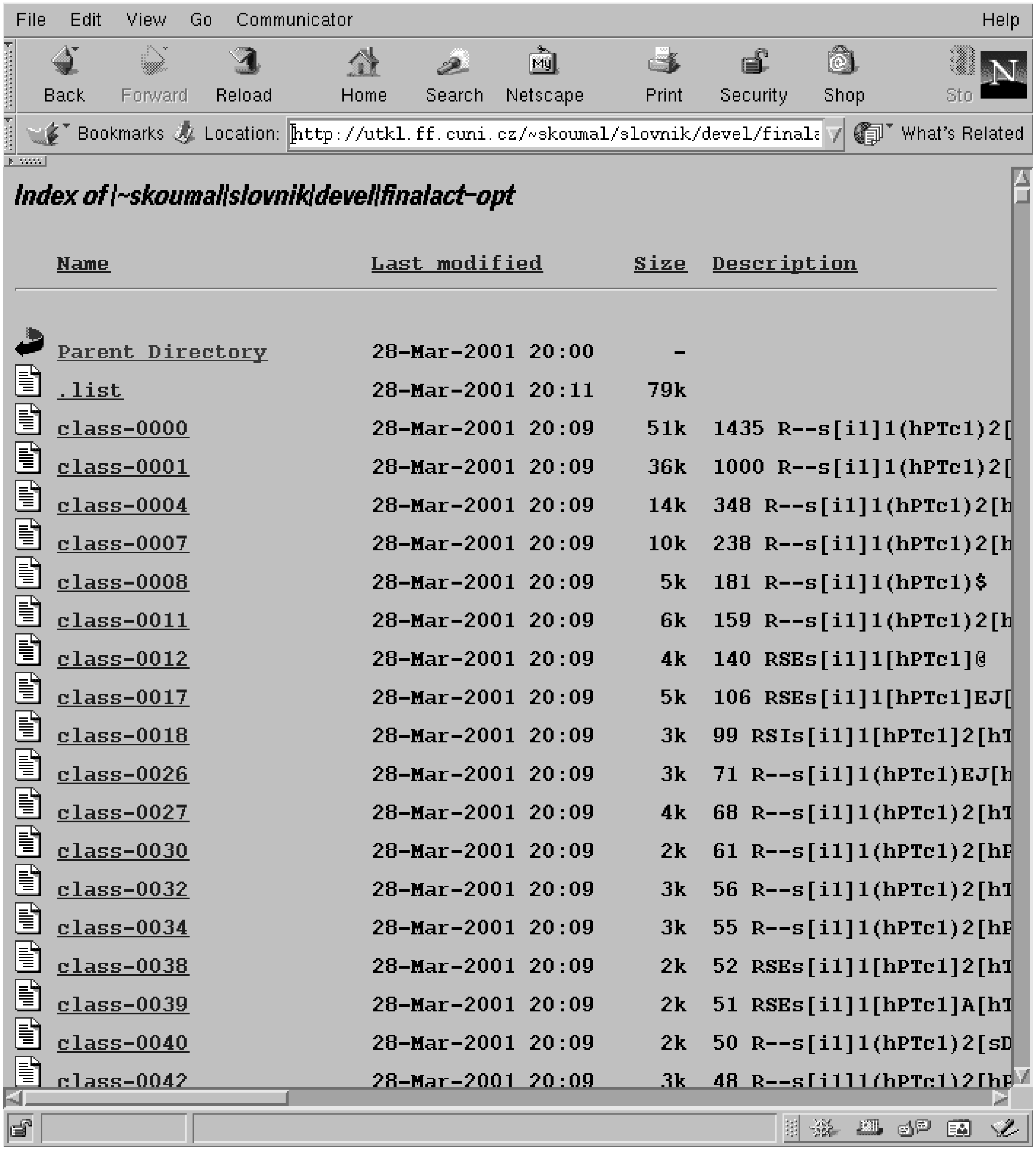


Figure G.3.: Frames processed fully automatically, with ambiguous free modifications